



TITLE:

X-ray crystallographic studies on catalytic mechanism and substrate recognition of bacterial unsaturated glycoside hydrolases(Dissertation_全文)

AUTHOR(S):

Itoh, Takafumi

CITATION:

Itoh, Takafumi. X-ray crystallographic studies on catalytic mechanism and substrate recognition of bacterial unsaturated glycoside hydrolases. 京都大学, 2006, 博士(農学)

ISSUE DATE:

2006-11-24

URL:

<https://doi.org/10.14989/doctor.r11948>

RIGHT:

**X-ray Crystallographic Studies on Catalytic Mechanism and
Substrate Recognition of Bacterial Unsaturated Glycoside Hydrolases**

Takafumi ITOH

2006

**X-ray Crystallographic Studies on Catalytic Mechanism and
Substrate Recognition of Bacterial Unsaturated Glycoside Hydrolases**

Takafumi ITOH

2006

CONTENTS

INTRODUCTION	1
CHAPTER I	CRYSTAL STRUCTURE OF A NOVEL UNSATURATED GLUCURONYL HYDROLASE	5
CHAPTER II	A NOVEL GLYCOSIDE HYDROLASE FAMILY 105	20
CHAPTER III	A NOVEL ENZYME REACTION CATALYZED BY UNSATURATED GLYCOSIDE HYDROLASES	36
Section 1	Structure of Unsaturated Glucuronyl Hydrolase Complexed with Substrate	36
Section 2	Substrate Recognition of Unsaturated Glucuronyl Hydrolase	54
Section 3	Structure of Unsaturated Galacturonyl Hydrolase Complexed with Substrate	64
CONCLUSION	76
ACKNOWLEDGEMENTS	77
LIST OF PUBLICATIONS	78

INTRODUCTION

Glycosaminoglycans such as chondroitin, hyaluronan, and heparin are highly negatively charged polysaccharides with a repeating disaccharide unit consisting of an uronic acid residue (glucuronic or iduronic acid) and an amino sugar residue (glucosamine or galactosamine) (Iozzo, 1998). Chondroitin consists of D-glucuronic acid (GlcA) and *N*-acetyl-D-galactosamine (GalNAc) along with a sulfate group(s) at position 4 or 6 or at both positions. Hyaluronan, which is also a member of the glycosaminoglycan family, is composed of GlcA and *N*-acetyl-D-glucosamine (GlcNAc). These major components of the mammalian extracellular matrix are responsible for cell-to-cell association, and are widely present in human tissues such as the eye, brain, liver, skin, and blood (Lindahl & Hook, 1978).

Glycosaminoglycans in the extracellular matrix are enzymatically degraded by hydrolases and lyases (Ernst *et al.*, 1995). Generally, hydrolases, which cleave bonds between glycosidic oxygen and anomeric carbon atoms through the addition of a water molecule, play an important role in the metabolism of glycosaminoglycans in mammals. Microbial pathogens that invade host cells degrade glycosaminoglycans primarily by the action of lyases (Hascall *et al.*, 1995; Linker *et al.*, 1955). Microbial lyases recognize the uronic acid residue in polysaccharides and produce unsaturated disaccharides with the uronic acid residue having a C=C double bond at the nonreducing terminus through the β -elimination reaction. Typical pathogens that degrade glycosaminoglycans include streptococci such as *Streptococcus agalactiae*, *Streptococcus pneumoniae*, and *Streptococcus pyogenes*. Streptococci cause severe infectious diseases, e.g., pneumonia, bacteremia, sinusitis, and meningitis; these bacteria degrade glycosaminoglycans by using polysaccharide lyases, which function as virulence factors (Jedrzejewski, 2001).

Unsaturated glucuronyl hydrolase (UGL), a member of glycoside hydrolase family GH-88 in the CAZy database (Henrissat, B., Coutinho, P., and Deleury, E., 1999;

<http://afmb.cnrs-mrs.fr/~cazy/CAZY/index.html>), acts on unsaturated oligosaccharides that have an unsaturated GlcA (Δ GlcA) with a α -glycoside bond, i.e., Δ GlcA-GalNAc produced by chondroitin lyase, Δ GlcA-GlcNAc produced by hyaluronate lyase, Δ GlcA-mannose (Man)-glucose (Glc) produced by xanthan lyase, and Δ GlcA-Glc-rhamnose (Rha)-Glc produced by gellan lyase (Hashimoto *et al.*, 1999) (Fig. 1). The enzyme probably produces Δ GlcA and the leaving saccharide from these unsaturated oligosaccharides through the hydrolytic reaction. The resultant product, Δ GlcA, is thought to be spontaneously converted to the open chain form because the ringed form of Δ GlcA is not obtained due to keto-enole equilibrium (Linker *et al.*, 1955; Hashimoto *et al.*, 1999). UGL and its gene were first identified in *Bacillus* sp. GL1 and were considered to be involved in the degradation of mammalian glycosaminoglycans and bacterial biofilms (*Sphingomonas* gellan and *Xanthomonas* xanthan) (Hashimoto *et al.*, 1999). Proteins showing high homology with UGL of *Bacillus* sp. GL1 were subsequently found to be encoded in the genomes of pathogenic bacteria, i.e., bacteroides, enterococci, mycoplasma, streptococci, and vibrio (Fig. 2). The enzyme thus appears to be responsible for the complete degradation of glycosaminoglycans. In fact, streptococcal hypothetical proteins homologous to UGL were recently found to exhibit enzyme activity (Hashimoto *et al.*, in preparation).

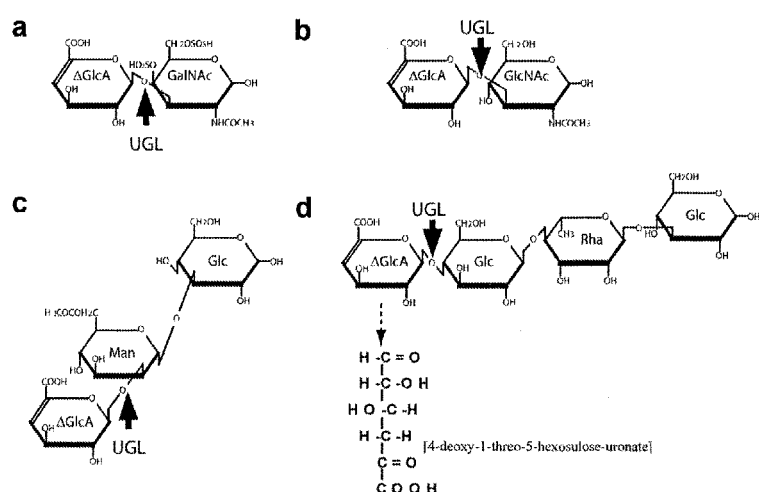


Fig. 1. Structures of unsaturated oligosaccharides. (a) Unsaturated chondroitin disaccharide (Δ GlcA-GalNAc), (b) unsaturated hyaluronate disaccharide (Δ GlcA-GlcNAc), (c) unsaturated xanthan trisaccharide (Δ GlcA-Man-Glc), and (d) unsaturated gellan tetrasaccharide (Δ GlcA-Glc-Rha-Glc). Arrows indicate the cleavage site of UGL. The dashed-line arrow indicates UGL hydrolysis.

```

      20      40      60      80
B. sp. GL1 -----HWQQAIGDALGITARNLKKFGDRFPHVSDGSE
B. halodurans -----MMWQAMTDVAEKLTLNKKRFGGRFPHVSEDE
S. pyogenes MARPLKTIALEPIKQFERFTKEDEPLSQEDITQALCLALKQVRLAMDYFYEDFTPTATKDN
S. pneumoniae ---HIKVTIEIKKSPERFLVPLLTKEEVGGQAIUKVIRQLELALDYFEDFTPTATFDM
                        **      *      *      *

      100     120     140     160
B. sp. GL1 KYVLNDHTQWTDGFWSGILWLCYBYTGDQCYREGAVRTVASFBRLEDFENLDHHDIGFL
B. halodurans HYELANNNEWTNGFWSGILWLCYBYTNDPAFRQAASTVRSFQGRMEONLELDHHDIGFL
S. pyogenes QYATMDNTETWTFWFCGLWLAYEYSGDDAIKALAQANDLSPLLRVTRDIELDHDHIGFL
S. pneumoniae VYPIMDNTETWTFWFGWELWLAYEYSGQDAFKNTIAHKQVLSPLDRVNRKVELDHDHIGFL
      *      *      *      *      *      *      *      *      *      *      *      *

      180     200     220     240
B. sp. GL1 YLSLSAKAQNIWEKDESAKRLALDAADVLMRRNRADAGIIQAWGPKGLPENGGRIIIDCLL
B. halodurans YLSLSAKAQNIWERDERAKQLTTEAADVLMRRNRREKIRLFOAWGPPGDLSSNGRIIVDCLM
S. pyogenes YTPSCMAENKLLKTPESREAAKKAADKLVCRYQDKGSPQAWGELGKED-YRLIIDCLL
S. pneumoniae YTPSCMAEYKINGDGEAREATLKAADKLIRYQKGGFTQAWGDLGKKEH-YRLIIDCLL
      *      *      *      *      *      *      *      *      *      *      *      *

      260     280     300     320
B. sp. GL1 NLPLLLWAGEQTGDPEYRKVAEAAHLKSRHFLVRGDDSSYHTFYFDPENGINAIGGTHQG
B. halodurans NLPLLFWASEVTGNPDYRRAATIADKTRRFRIVRGDDSTVHTFYFENQCTGEALGGTHQG
S. pyogenes NIQLLFWASQETGDNRYRDMANHFYASANNVIRDDASAYHTFYFDPETGDPVKGVTGQG
S. pneumoniae NIQLLFFAYGETGDKYDYDIAESEFYASANNVIRDDASSEHTFYFDPETGQPFKGVTRQG
      *      *      *      *      *      *      *      *      *      *      *      *

      340     360     380     400
B. sp. GL1 NTDGSITWTRGQAWGIYGFALNSRYLGNADLLETAKRMARHFLARVPEDGVVYWDPEVQFE
B. halodurans YEDGSIWSSGQAWAIYGFALNYRYTGNERYLETAKRTAKYFISGLPADYVAYWDFNAPIT
S. pyogenes YSDGSWARGQAWGIYGLPLTYRFLKEPELIQLEKGMTHYFLNELPKDQVSYWDLIFGDSG
S. pneumoniae YSDGSWARGQSWGYGIPLTYRFLKESCFDLFGGVTHYFLNELPKDHVEYWDLIFNDG
      *      *      *      *      *      *      *      *      *      *      *      *

      420     440     460     480
B. sp. GL1 PSSYRDSASAIACGLLEIASQLDESDEPERORFIDAAKTTVTALRGYAEERDDGEAEGF
B. halodurans FQTRKDSASAIASGILELLSHLQETDEPKAFQSQSVQKQMTLVENYASEKD-AQGL
S. pyogenes SEQSRSASATAIIVCGIHEMLKTLPOHDPEKTTYEAHHSMRLALIKDYANKDLPGAPL
S. pneumoniae SDQSPDSASATAIIVCGIHEMLKHLPEVDALKDIYKHANHAHLRLIENYANDQETPGGTS
      *      *      *      *      *      *      *      *      *      *      *      *

      500     520     540     560
B. sp. GL1 IIRGGSYHVVRGGISFDDYTINGDYTYLFAALIRLERGVIGYNYERGR
B. halodurans IIRGGSYSVRIGHAFDDYTYNGDYTYLFAALIRLERGVIGYNYERGR
S. pyogenes LLHGVSYSWASGKGVDEGNIWGDYTYLFAALIRFYKDNPNY-----
S. pneumoniae LLHGVSYSWASGKGVDEGNIWGDYTYLFAALIRFYKDNHLY-----
      *      *      *      *      *      *      *      *      *      *      *

```

Fig. 2. Amino acid sequence alignment of UGL and other homologous bacterial proteins made using the ClustalW program (<http://clustalw.genome.ad.jp/>). *B. sp. GL1*, UGL of *Bacillus sp. GL1* (Ac. No., AB019619-2); *B. halodurans*, protein of *Bacillus halodurans* (Ac. No., AP001514-60); *S. pyogenes*, protein of *Streptococcus pyogenes* (Ac. No., AE006517-10); and *S. pneumoniae*, protein of *Streptococcus pneumoniae* (Ac. No., AE007344-9). Identical amino acid residues among the four proteins are indicated by asterisks.

Structural analysis of lyases and UGL is indispensable for clarifying molecular mechanisms underlying the recognition and degradation of unsaturated saccharides and sequential reaction mechanisms involved in polysaccharide depolymerization by bacteria. The crystal structures of polysaccharide lyases acting on chondroitin AC (Féthière *et al.*, 1999; Lunin *et al.*, 2004), chondroitin B (Huang *et al.*, 1999), chondroitin ABC (Huang *et al.*, 2003), hyaluronate (Li *et al.*, 2000; Li & Jedrzejewski, 2001; Jedrzejewski *et al.*, 2002; Smith *et al.*, 2005), and xanthan (Hashimoto *et al.*, 2003; Maruyama *et al.*, 2005) have been determined, and the structure-function relationships of these lyases have been studied.

In this thesis, in order to clarify the structure-function relationship of unsaturated glycoside hydrolases (UGL from *Bacillus sp. GL1* and its tertiary structural homologue YteR from *Bacillus subtilis*), the determination of the X-ray crystal structure of UGL (CHAPTER I), the identification of YteR as unsaturated galacturonidase (CHAPTER II), and the elucidation of the catalytic

reaction mechanism of these unsaturated glycoside hydrolases (CHAPTER III) were carried out. These results provide useful information on the molecular design of drugs for the treatment of infectious diseases in mammalian and plant cells.

References

- Ernst, S., Langer, R., Cooney, C. L., and Sasisekharan, R. (1995) *Crit. Rev. Biochem. Mol. Biol.* **30**, 387-444
- Féthière, J., Eggimann, B., and Cygler, M. (1999) *J. Mol. Biol.* **288**, 635-647
- Hascall, V. C., Midura, R. J., Sorrell, J. M., and Plaas, A. H. (1995) *Adv. Exp. Med. Biol.* **376**, 205-216
- Hashimoto, W., Kobayashi, E., Nankai, H., Sato, N., Miya, T., Kawai, S., and Murata, K. (1999) *Arch. Biochem. Biophys.* **368**, 367-374
- Hashimoto, W., Nankai, H., Mikami, B., and Murata, K. (2003) *J. Biol. Chem.* **278**, 7663-7673
- Huang, W., Matte, A., Li, Y., Kim, Y. S., Linhardt, R. J., Su, H., and Cygler, M. (1999) *J. Mol. Biol.* **294**, 1257-1269
- Huang, W., Lunin, V. V., Li, Y., Suzuki, S., Sugiura, N., Miyazono, H., and Cygler, M. (2003) *J. Mol. Biol.* **328**, 623-634
- Iozzo, R. V. (1998) *Annu. Rev. Biochem.* **67**, 609-652
- Jedrzejewski, M. J. (2001) *Microbiol. Mol. Biol. Rev.* **65**, 187-207
- Jedrzejewski, M. J., Mello, L. V., De Groot, B. L., and Li, S. (2002) *J. Biol. Chem.* **277**, 28287-28297
- Li, S., Kelly, S. J., Lamani, E., Ferraroni, M., and Jedrzejewski, M. J. (2000) *EMBO J.* **19**, 1228-1240
- Li, S., and Jedrzejewski, M. J. (2001) *J. Biol. Chem.* **276**, 41407-41416
- Lindahl, U., and Hook, M. (1978) *Annu. Rev. Biochem.* **47**, 385-417
- Linker, A., Meyer, K., and Weissmann, B. (1955) *J. Biol. Chem.* **213**, 237-248
- Lunin, V. V., Li, Y., Linhardt, R. J., Miyazono, H., Kyogashima, M., Kaneko, T., Bell, A. W., and Cygler, M. (2004) *J. Mol. Biol.* **337**, 367-386
- Maruyama, Y., Hashimoto, W., Mikami, B., and Murata, K. (2005) *J. Mol. Biol.* **350**, 974-986
- Smith, N. L., Taylor, E. J., Lindsay, A. M., Charnock, S. J., Turkenburg, J. P., Dodson, E. J., Davies, G. J., and Black, G. W. (2005) *Proc. Natl. Acad. Sci. USA* **102**, 17652-17657

CHAPTER I

CRYSTAL STRUCTURE OF A NOVEL UNSATURATED GLUCURONYL HYDROLASE

Together with polysaccharide lyases, bacterial UGL plays an important role in depolymerizing polysaccharides constituting mammalian extracellular matrix and/or bacterial biofilm. A large number of UGL homologues are encoded in the genome of pathogenic bacteria such as streptococci and vibrio, suggesting that these are crucial for invasion of bacterial cells into host cells through the depolymerization of the extracellular matrix. Structural and functional analyses of UGL as well as polysaccharide lyases are therefore important for the clarification of the ecosystem of pathogenic and/or biofilm-dependent bacteria and for the establishment of a novel therapy for bacterial infectious diseases. Although many polysaccharide lyases, e.g. chondroitin (Féthière *et al.*, 1999; Lunin *et al.*, 2004; Huang *et al.*, 1999, 2003), heparin (Shaya *et al.*, 2006), hyaluronate (Li *et al.*, 2000; Li & Jedrzejewski, 2001; Jedrzejewski *et al.*, 2002; Smith *et al.*, 2005), and xanthan lyases (Hashimoto *et al.*, 2003; Maruyama *et al.*, 2005), have functionally and structurally been well characterized, little information on structure-function relationship of UGL has been accumulated.

In order to obtain clues to clarify structural characteristics, this chapter deals with the determination of crystal structure of *Bacillus* sp. GL1 UGL by X-ray crystallography.

Materials and Methods

Crystallization and X-ray diffraction

UGL of *Bacillus* sp. GL1 was overexpressed in *Escherichia coli*, purified, and crystallized by sitting-drop vapor diffusion as described previously (Mori *et al.*, 2003a, b). UGL crystals were soaked in a heavy-atom derivative solution containing 1 mM K_2PtCl_4 , 1 mM $Hg(CH_3COO)_2$, or 0.5

mM NaAuCl₄ for 15-50 minutes at 20°C. These heavy-atom solutions were prepared in 52% (v/v) 2-methyl-2,4-pentanediol (MPD), 0.12 M sodium chloride, 0.1 M glycine, and 0.1 M Tris-HCl (pH 7.6). The crystals were removed from a droplet on a mounted nylon loop (Hampton Research, Laguna Niguel, CA, USA), and placed in a cold nitrogen gas stream at -173°C. X-ray diffraction images of the UGL crystal (Native 1) were collected using a Quantum 4R CCD area detector (ADSC, CA, USA) with synchrotron radiation at a wavelength of 0.72 Å at the BL-38B1 station of SPring-8 (Japan). Images were processed with DENZO and SCALEPACK software (Otwinowski & Minor, 1997) to a resolution of 1.8 Å (Table 1). Diffraction images of another crystal (Native 2) and derivative crystals for phasing were collected with a Bruker Hi-Star multiwire area detector using CuK α radiation generated by a MAC Science M18XHF rotating anode generator, and were processed with SADIE and SAINT software (Bruker, Karlsruhe, Germany).

Table 1. Synchrotron radiation data collection and refinement statistics for a UGL crystal (Native 1)

Crystal system	Hexagonal
Space group	<i>P</i> 6 ₂ 22
Unit cell parameters (Å)	a = b = 103.01, c = 223.04
Molecules/Asym. Unit	1
Data collection	
Resolution limit (fast shell) ^a (Å)	25.0 – 1.80 (1.86 – 1.80)
Measured reflections	300292 (24083)
Unique reflections	63892 (6175)
Redundancy	4.7 (3.9)
Completeness ($\sum I > \sigma(I)$) (%)	97.6 (96.0)
R_{merge} (%) ^b	9.6 (29.5)
Refinement	
Final model	377 residues, 478 water, 4 glycine, 2 DTT, and 1 MPD
Resolution limit (Å)	25.0 – 1.80 (1.91 – 1.80)
Used reflections	63316 (9024)
Completeness ($\sum F > 2\sigma(F)$) (%)	96.7 (93.9)
Average <i>B</i> -factor (Å ²)	16.3
<i>R</i> -factor (%) ^c	16.8 (22.4)
R_{free} (%) ^d	18.9 (24.6)
rms deviations	
Bond (Å)	1.22
Angle (deg)	0.005

^a Data in highest resolution shells is given in parentheses.

^b $R_{\text{merge}} = \sum |I_i - \langle I \rangle| / \sum I_i \times 100$, where I_i is the intensity of individual reflection and $\langle I \rangle$ the mean intensity of all reflections.

^c $R\text{-factor} = \sum |F_o - F_c| / \sum |F_o| \times 100$, where F_o is the observed structure factor and F_c the calculated structure factor.

^d R_{free} was calculated from a randomly chosen 10% of reflections as defined by the CNS.

Table 2. Phasing statistics

Compound	Native 2	K ₂ PtCl ₄	Hg(CH ₃ COO) ₂	NaAuCl ₄
Concentration (mM) ^a		1.0	1.0	0.5
Soaking time (min)		50	30	15
Resolution limit (Å)	2.7	3.5	3.5	3.5
Phasing power		1.08	1.23	1.06
R_{calc}		0.612	0.652	0.701
R_{omit}		0.088	0.211	0.104
Number of sites		2	2	2
Binding sites		Met-1, Met-121	Cys-150, Cys-287	Cys-150, His-210

^a Soaking solution consisted of 52% (v/v) MPD, 0.12 M of sodium chloride, 0.1 M of glycine, and 0.1 M of Tris - HCl buffer (pH 7.6).

Mean figure of merit: 0.426.

Structure determination and refinement

The UGL crystal structure was determined by multiple isomorphous replacement (mir). Phase calculation and refinement were done with Native 2 and derivative data sets using a PHASES program (Furey & Swaminathan, 1997). Major sites of heavy atoms were determined by the interpretation of difference Patterson maps calculated at a resolution of 6.0-3.5 Å. Additional heavy-atom sites were determined from difference Fourier maps. Phasing results are listed in Table 2. The mean figure of merit was 0.426. The phase was greatly improved and the mean figure of merit increased to 0.754 after solvent flattening (Wang, 1985) with PHASES program. Initial model building was done with Native 2 data sets and the phase at 3.5 Å using the TURBO-FRODO program (AFMB-CNRS, Marseille, France) on a Silicon Graphics Octane computer. Simulated annealing refinement was done with this model and 25-2.5 Å resolution data from Native 1 data sets with CNS ver. 1.1 (Brünger *et al.*, 1998). The model was heated to 3,000 K, then slowly cooled to 300 K (time step, 0.5 fs; decrease in temperature, 25 K; number of steps at each temperature, 50), and 200 cycles of Powell minimization were done. $F_o - F_c$ and $2F_o - F_c$ maps were used to locate the correct model. Several rounds of positional and *B*-factor refinement, followed by manual model building, were done to improve the model by increasing data to a resolution of 1.8 Å. Water molecules were incorporated where the difference density exceeded the mean by 3.0 σ or more and the $2F_o - F_c$ map showed a density exceeding 1.0 σ . Seven fragments of nonprotein or nonwater density were modeled into four glycine molecules, two dithiothreitol (DTT) molecules, and one MPD molecule from the crystallization medium, and density was excellent for the whole molecule. The final *R*-factor was 16.8% for 63,316 data points with $F > 2.0 \sigma (F)$ in a resolution of 25.0-1.8 Å (96.7% completeness). The *R*-free value calculated for randomly separated 10% data was 18.9%. The stereoquality of the model was assessed using the PROCHECK (Laskowski *et al.*, 1993) and WHAT-CHECK (Hooft *et al.*, 1996) programs. Structural similarity was searched for in the RCSB Protein Data Bank (Berman *et al.*, 2000) using the DALI program (Holm & Sander, 1993).

Coordinates of hypothetical protein YteR (INC5) were taken from the RCSB Protein Data Bank. UGL and YteR models were superimposed by a fitting program in TURBO-FRODO. Ribbon plots were prepared using the MOLSCRIPT (Kraulis, 1991), Raster3D (Merrit & Murphy, 1994), and GRASP (Nicholls *et al.*, 1991) programs.

Mutagenesis and CD spectra measurement

Asp-88 or Asp-149 in UGL was replaced with an asparagine residue by the use of a QuickChange site-directed mutagenesis kit (Stratagene) and mutation was confirmed by DNA sequencing. The plasmid pET3a-UGL, which is the expression vector for wild-type UGL (Mori *et al.*, 2003b), was used as a template. Primers were as follows: D88N (Asp-88 → Asn-88), 5'-AGAATCTCGATCATCACAACATCGGCTTCCTATACTC-3' and 5'-GAGTATAGGAAGCCGAT**GTT**GTGATGATCGAGATTCT-3'; and D149N (Asp-149 → Asn-149), 5'-GGACGCATCATCATCA**ACT**GCCTGCTGAATCTG-3' and 5'-CAGATTCAGCAGGCAGTTGATGATGATGCGTCC-3' (mutations are indicated by bold letters). The *E. coli* BL21(DE3) was transformed with plasmids having a mutation, i.e., pET3a-UGL(D88N) and pET3a-UGL(D149N). Mutant enzymes were expressed and purified as for wild-type UGL (Mori *et al.*, 2003b). Structural conformations of purified wild-type and mutant enzymes were evaluated by far-UV circular dichroism (CD) spectroscopy using a Jasco J720 spectropolarimeter at 190-260 nm with a demountable quartz cell having a 0.1 mm path length.

Enzyme assay

UGL reactions for the wild-type and mutants were conducted at 30°C as follows: The reaction mixture consisted of 50 mM sodium phosphate (pH 6.5), 20 – 500 µM substrate, and enzymes in a 500 µl reaction volume. Enzyme activity was measured by monitoring the decrease in absorbance at 235 nm, corresponding to the loss of the C=C double bond of the substrate because the pyranose ring of the released ΔGlcA readily opens so that it is nonenzymatically converted to α-keto acid through the loss of the double bond (Fig. 1d, Introduction) (Linker *et al.*, 1955;

Hashimoto *et al.*, 1999). Enzyme concentration was determined by UV spectrophotometry using theoretical molar extinction coefficient $\epsilon_{280} = 99,570 \text{ (M}^{-1}\text{cm}^{-1})$. Enzyme purity was assessed by SDS-PAGE followed by Coomassie Brilliant Blue staining. The gellan lyase product ($\Delta\text{GlcA-Glc-Rha-Glc}$) for the UGL substrate was prepared as described elsewhere (Hashimoto *et al.*, 1997). Hyaluronate lyase product ($\Delta\text{GlcA-GlcNAc}$) and chondroitin lyase product ($\Delta\text{GlcA-GalNAc}$) were obtained from Seikagaku Corporation (Tokyo, Japan). k_{cat} and K_{m} parameters were determined by nonlinear fitting to the Michaelis-Menten equation.

Results and Discussion

Crystallization and structure determination

UGL of *Bacillus* sp. GL1 is a monomeric enzyme with a molecular mass of about 43 kDa (377 amino acid residues) (Hashimoto *et al.*, 1999). A UGL crystal ($0.1 \times 0.1 \times 0.5 \text{ mm}$) was obtained by sitting-drop vapor diffusion as described previously (Mori *et al.*, 2003a). The space group was determined to be $P6_522$ with unit cell dimensions of $a = b = 103.01$ and $c = 223.04 \text{ \AA}$, and the solvent content was 69% assuming one molecule per asymmetric unit. Results of native data collection using synchrotron radiation at the BL-38B1 station of SPring-8 are summarized in Table 1. The phase of the structure was solved by mir. Table 2 shows phasing statistics at a resolution of 3.5 \AA .

Quality of refined model

The refined model of UGL consists of 377 amino acid residues, and 478 water molecules, four glycine molecules, two DTT molecules, and one MPD molecule. The entire polypeptide chain sequence was well traced, and the electron densities of the main-chain and side-chain were generally very well defined in the $2F_o - F_c$ map, except for N-terminal amino acid residue Met-1 and C-terminal amino acid residue Arg-377, whose electron density was too low for them to be identified completely. Other ligand molecules were also well fitted. The final overall R -factor for

the refined model was 16.8%, with 63,316 unique reflections at a resolution of 25.0-1.8 Å. The final free *R*-factor was 18.9%. Final root-mean-square (rms) deviations from standard geometry were 0.005 Å for bond lengths and 1.22° for bond angles. Based on theoretical curves in the plot calculated according to Luzzati (1952), the absolute positional error was estimated to be close to 0.17 Å at a resolution of 5.0-1.8 Å. Most (88.3%) nonglycine residues lie within most favored regions, and other residues (11.3%) within additionally allowed regions of the Ramachandran plot as defined in PROCHECK (Laskowski *et al*, 1993). Ser-345 ($\phi = 38^\circ$, $\psi = 65^\circ$), however, falls in generously allowed regions, exhibiting well-defined density in the $2Fo - Fc$ map and located in a sharp bend of the loop neighboring a helix. There are no residues in disallowed regions.

Overall UGL structure

The overall structure of UGL is shown as ribbon models (Fig. 1a, b) and a molecular surface model (Fig. 1c). The enzyme is approximately 45× 45× 40 Å and consists of an α_6/α_6 -barrel structure with a deep pocket that is likely to be the active site (described below). Fig. 2 shows the UGL structure's topology. The structure consists of 12 long α -helices (H1, 3-20; H2, 44-58; H3, 61-79; H4a, 89-94; H4b, 96-104; H5, 107-120; H6, 154-164; H7, 167-183; H8, 220-237; H9, 240-254; H10, 279-294; H11, 301-320; and H12, 351-366), 3 antiparallel β -sheets consisting of 2-3 strands and designated SA to SC (SA1, 24-28; SA2, 34-37; SB1, 124-125; SB2, 130-131; SC1, 145-147; SC2, 194-197; and SC3, 204-208), 2 short 3_{10} -helices (121-123 and 148-153) adjoining the long α -helices, and some loops. H4 has one additional residue, Leu-95, whose oxygen atom has no hydrogen bond with the nitrogen atom of a paired residue. This α -helix (H4) was thus divided into two segments, H4a and H4b. The α -helix bends at this point, the angle being 30° between H4a and H4b. The α_6/α_6 -barrel structure is formed by six outer helices (H1, H3, H5, H7, H9, and H11) running in roughly the same direction and six inner helices (H2, H4, H6, H8, H10, and H12) oriented in the opposite direction. These helices are connected, in a nearest-neighbor and an up-and-down pattern by short and long loops. Loops between α -helices, such as H1 and H2, are

referred to as L-H1:H2. One side of the barrel has short loops (L-H2:H3, L-H4:H5, L-H6:H7, L-H8:H9, 2 residues; and L-H10:H11, 6 residues). The other side, including the pocket, consists of long loops (L-H1:H2, 23 residues; L-H3:H4, 9 residues; L-H5:H6, 33 residues; L-H7:H8, 36 residues; L-H9:H10, 24 residues; and L-H11:H12, 30 residues). These long loops are packed together and form the wall of the deep funnel-shaped pocket roughly 15-20 Å in diameter at the lip and about 15 Å deep (Fig. 1c). This pocket is widely surrounded by aromatic residues.

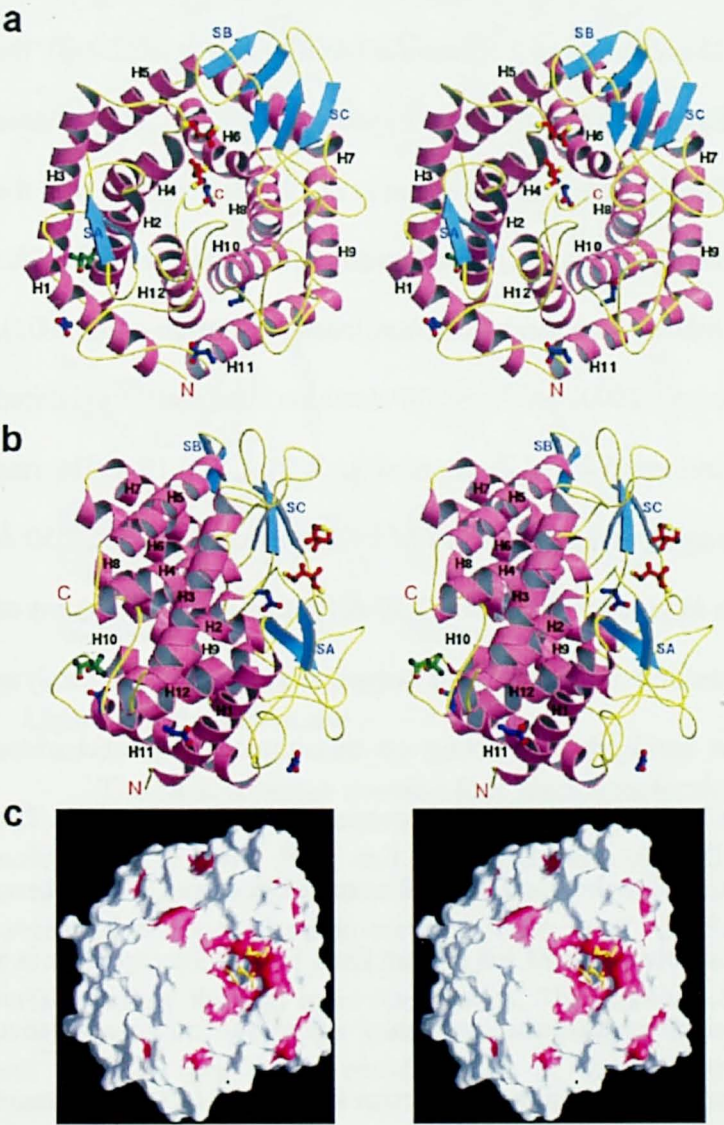
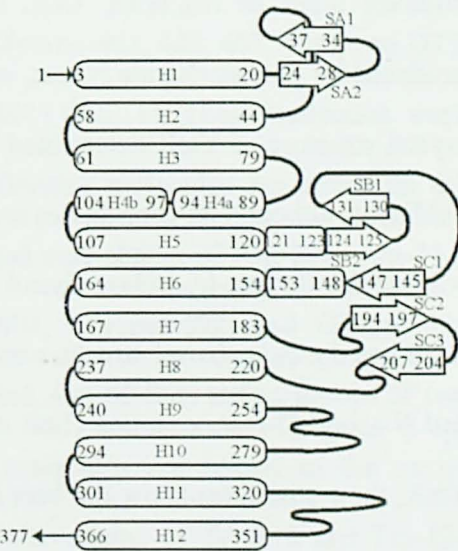


Fig. 1. Overall structure of UGL bound with molecules. (a), (b) Colors denote secondary structure elements (pink, α -helices; cyan, β -strands; yellow, loops and coils), glycine (blue), DTT (red), and MPD (green) molecules (a, front; b, side). (c) The structure is represented as a white molecular surface model. Aromatic residues are pink. Bound molecules are presented by bond models and colored yellow.

Fig. 2. Topology of the fold in UGL. UGL includes an α_6/α_6 -barrel. The 12 α -helices in the α_6/α_6 -barrel are numbered from the N-terminal, H1 to H12, as in Fig. 1. Six boxes (H1, H3, H5, H7, H9, and H11) are outer α -helices facing the solvent. The other six boxes (H2, H4, H6, H8, H10, and H12) are inner α -helices. The 3_{10} -helix is also shown as boxes. The three short, antiparallel β -sheets, designated SA to SC, are shown as arrows.



Structural comparison

UGL consists of an α/α -toroidal fold. This basic fold is common in glycoside hydrolases, polysaccharide lyases, and terpenoid cyclases/protein prenyltransferases in the SCOP database (<http://scop.mrc-lmb.cam.ac.uk/scop/index.html>) (Murzin *et al.*, 1995). UGL has the α_6/α_6 -barrel found in the six-hairpin enzyme superfamily of the SCOP database, which includes glucoamylases (Aleshin *et al.*, 1992, 2003; Ševčík *et al.*, 1998), cellulase catalytic domains (Juy *et al.*, 1992; Alzari *et al.*, 1996; Sakon *et al.*, 1997), *N*-acetyl-D-glucosamine 2-epimerase (AGE) (Itoh *et al.*, 2000), the maltose phosphorylase central domain (Egloff *et al.*, 2001), and hypothetical protein YteR. Based on the structural similarity in the RCSB Protein Data Bank (Berman *et al.*, 2000) observed with the DALI (Holm & Sander, 1993) program, three proteins, i.e., hypothetical protein YteR from *Bacillus subtilis*, glucoamylase from *Thermoanaerobacterium thermosaccharolyticum* (Aleshin *et al.*, 2003), and AGE from the porcine kidney (Itoh *et al.*, 2000), in the SCOP database superfamily exhibited the highest degree of similarity. These proteins exhibited Z-scores of 32.7, 21.1, and 19.4. The rms distance was 2.80 Å for the superimpositioning of 323 C α atoms of UGL on those of YteR, 3.50 Å for that of 287 C α atoms of UGL on those of glucoamylase, and 3.30 Å for that of 278 C α atoms of UGL on those of AGE, although they exhibit no amino acid sequence similarity and catalyze different types of reactions. UGL is an exo-hydrolase acting on unsaturated oligosaccharides produced by polysaccharide lyases, while YteR is a hypothetical protein of unknown function. The crystal structure of YteR determined by the Midwest Center for Structural Genomics has not been published, although its coordinates are available in the RCSB Protein Data Bank. Glucoamylase is a polysaccharide exo-hydrolase found in some prokaryotic and many eukaryotic microorganisms. AGE has the α_6/α_6 -barrel structure and catalyzes the epimeric reaction for *N*-acetyl-D-glucosamine and *N*-acetyl-D-mannosamine (Itoh *et al.*, 2000). Fig. 3 shows the superimpositioning of UGL on YteR; their structures show the best fit overall. The 12 helices of the α_6/α_6 -barrel are very similar for UGL and YteR in position, direction, and angle, but significant differences exist between UGL

and YteR in the loop structure. The L-H1:H2 loop of UGL is longer (23 residues) than that of YteR (13 residues). The L-H7:H8 loop of UGL protrudes from the wall to the outside, while that of YteR is directed into the pocket. L-H11:H12 of UGL makes the pocket open wide, while that of YteR makes the pocket close.

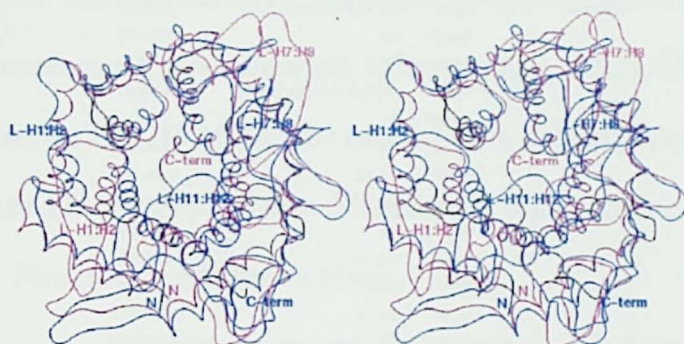


Fig. 3. Structural comparison of overall structures of UGL (pink) and hypothetical protein YteR (blue). Superimposed results are shown schematically in α -traces. Coordinates of YteR (INC5) were taken from the RCSB Protein Data Bank.

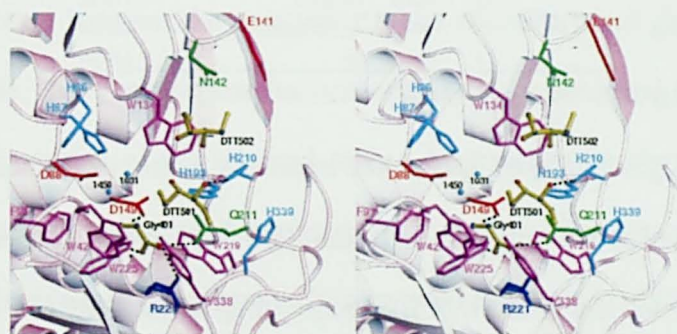


Fig. 4. Stereoview of the active site in the UGL pocket. Amino acid residues face the putative active site, glycine, and DTT molecules. Side-chains are red (Asp), blue (Arg), purple (Trp, Tyr, Phe), cyan (His), and green (Gln). Glycine and DTT molecules are yellow. Hydrogen bonds are shown as dotted lines.

Ligands in the active site

The UGL structure contains four glycine molecules (Glycine-401, 402, 403, 404), two DTT molecules (DTT-501, 502), and one MPD molecule (MPD-601) (Fig. 1). These molecules were present in the crystallization medium (Mori *et al.*, 2003a). Glycine molecules are bound on the surface mainly through ionic interactions. The positive-charged side-chains of Arg-221, Arg-334, and Arg-303 face the carboxyl groups of three Glycine-401, Glycine-402, and Glycine-404 molecules, and the negative-charged side-chains of Asp-149 and Asp-60 face amino groups of two Glycine-401 and Glycine-403 molecules. DTT and MPD molecules are bound in the cavity predominantly through a hydrophobic interaction. Aromatic side-chains of Tyr-338 and Trp-134 contact the DTT-501 and DTT-502 molecules, and Tyr-55 the MPD-601 molecule.

Notably, one glycine molecule (Glycine-401) and two DTT molecules (DTT-501, 502) are bound in the deep pocket, which was confirmed to be an active site in α_6/α_6 -barrel enzymes, and is formed by residues belonging to inner helices and long loops (Aleshin *et al.*, 1992, 1996, 2003; Ševčík *et al.*, 1998; Juy *et al.*, 1992; Alzari *et al.*, 1996; Sakon *et al.*, 1997; Itoh *et al.*, 2000; Egloff *et al.*, 2001; Guerin *et al.*, 2002; Parsiegla *et al.*, 2000; Van Petegem *et al.*, 2003) (Figs. 1c

and 4). The two O atoms of the carboxyl group of the Glycine-401 molecule, bound to the bottom surface of the pocket, are hydrogen-bonded to the Arg-221 N η 2 atom (2.9 Å), Gln-211 N ϵ 2 atom (3.0 Å), and Trp-225 N ϵ 1 atom (3.0 Å), and the N atom of its amino group to the Asp-149 O δ 1 atom (2.9 Å). For DTT molecules, in addition to hydrophobic interaction, the O2 hydroxyl group of DTT-501 is hydrogen-bonded to the His210 N ϵ 2 atom (3.0 Å) (Table 3). The inherent substrates, i.e., products of lyases (Fig. 1, Introduction), of UGL include carboxyl and hydroxyl groups, and pyranose rings of substrates appear to exhibit a stacked hydrophobic interaction with aromatic residues, as is often seen in complexes of polysaccharide lyases with substrates (Hashimoto *et al.*, 2003; Huang *et al.*, 1999, 2001; Mello *et al.*, 2002; Ponnuraj & Jedrzejewski, 2000). Glycine has a carboxyl group, and DTT contains two hydroxyl groups and a hydrophobic moiety. These structural characteristics of glycine and DTT molecules resemble those of substrates. The distance between Glycine-401 and DTT-502 is about 14 Å, which corresponds to the distance of a trisaccharide. Since UGL can act on unsaturated di-, tri-, and tetrasaccharides (Fig. 1, Introduction), these molecules

Table 3. Ligands and UGL interaction at the active site

Hydrogen bonds (< 3.2 Å)					
Molecule	Atom	Paired atom in UGL	Element	Distance (Å)	
Glycine-401	O1	Arg-221 Nη2	H8	2.9	
	O1	Gln-211 Nε2	H8	3.0	
	O2	Trp-225 Nε1	L-H7:H8	3.0	
	N	Asp-149 Oδ1	H6	2.9	
DTT-501	O2	His-210 Nε2	L-H7:H8	3.0	
van der Waals contact (C-C distance < 4.5 Å)					
Molecule	Atom	Paired atom in UGL	Element		
Glycine-401	C	Trp-42 Cε2, Cζ2, Cζ3, Cη2, Cζ2	L-H11:H12		
		Trp-219 Cζ2	L-H7:H8		
		Arg-221 Cζ	H8		
	Cα	Trp-42 Cδ2, Cε2, Cε3, Cζ2, Cζ3, Cη2	L-H11:H12		
		Asp-149 Cγ	H6		
DTT-501	C1	Trp-42 Cδ1	L-H11:H12		
		Tyr-338 Cε1, Cζ	L-H11:H12		
	C2	His-210 Cδ2	L-H7:H8		
	C3	His-210 Cδ2	L-H7:H8		
	C4	His-210 Cδ2, Cε2	L-H7:H8		
		Gln-211 Cγ, Cδ	L-H7:H8		
		Tyr-338 Cε1, Cζ	L-H11:H12		
		His-339 Cδ2	L-H11:H12		
DTT-502	C1	Trp-134 Cβ, Cγ	L-H5:H6		
	C2	Trp-134 Cβ, Cγ, Cδ2	L-H5:H6		
	C3	Trp-134 Cγ, Cδ2, Cα3, Cζ3	L-H5:H6		
	C4	His-210 Cε1	L-H7:H8		

(Glycine-401, DTT-501, and DTT-502) are thought to interact with UGL in the active site, and to be located at the three subsites for saccharide substrates. At the active site, side-chains of aromatic residues (Trp-42, Trp-134, Trp-219, Trp-225, Phe-91, and Tyr-338), positively-charged residues (Arg-221, His-86, His-87, and His-193), negatively-charged residues (Asp-88 and Asp-149), and a polar residue (Gln-211), which are completely conserved in UGL and its homologues (Fig. 2, Introduction), were observed. Other residues, whose side-chains also face the solvent at the active site, are His-210 of *Bacillus* sp. GL1 (Arg of other homologues), His-339 (Ser), Glu-141 (Ser), and Asn-142 (Asp, His). These differences in amino acid residues may imply substrate specificity.

Structural basis for catalysis

UGL releases nonreducing terminal Δ GlcA from unsaturated oligosaccharides by splitting the bond between the anomeric carbon of Δ GlcA and glycosidic oxygen (C1-O; α -1,3, α -1,2, or α -1,4). The reaction mechanism underlying unsaturated glucuronyl hydrolysis and the anomeric configuration of the released Δ GlcA have yet to be clarified. The pyranose ring of the released Δ GlcA, however, readily opens accompanying the loss of the double bond in C4=C5, and the saccharide is nonenzymatically converted to α -keto acid (Fig. 1d, Introduction) (Linker *et al.*, 1955; Hashimoto *et al.*, 1999). This makes it difficult to analyze the anomeric configuration in the UGL reaction. To verify implications about the active site of UGL above, and to determine the anomeric configuration of the intermediate monosaccharide, the preparation of enzyme crystals bound with sugars, substrates, or products by soaking or cocrystallization was attempted, but failed. The presence of ligand molecules (one glycine and two DTT) in the active pocket interfered with the binding of sugars. A catalytic mechanism was therefore proposed based on the UGL structure, conserved amino acid residues, and knowledge of glycoside hydrolysis, detailed below. Asp-88, Asp-149, His-86, His-87, His-193, and Arg-221, the ionizable residues nearest to the glycine molecule corresponding to the Δ GlcA binding site, are completely conserved in UGL and its homologues (Fig. 4 and; Fig. 2, Introduction), and may play an important role in glycoside

hydrolysis. The carboxyl group of Glycine-401 is believed to be located on that of Δ GlcA and has an ion-pair interaction with the side-chain of Arg-221, as described above. Arg-221 is therefore a residue important for Δ GlcA recognition. Two basic mechanisms, inversion and retention, have been proposed for glycoside hydrolases and are classified based on net retention or inversion of the anomeric configuration of the reaction product (Davies & Henrissat, 1995). Both reaction mechanisms involve two catalytic ionizable groups of carboxyl and carboxylate groups. Asp-88 and Asp-149 are thus the most likely catalytic residues for UGL. One is thought to act as the carboxylate group and nucleophile/base for anomeric carbon and the other as the carboxyl group and proton donor/acid for glycosidic oxygen in the first step in both reactions. Asp-88 and Asp-149 are 7.3 Å apart, a distance suitable for glycoside hydrolases that catalyze through inversion (average 10.0 Å) (McCarter & Withers, 1994). This inversion is generally accepted for almost all glycoside hydrolases with an α_6/α_6 -barrel, such as glucoamylases, endoglucanase CelD, CelA, and endo/exocellulase E4. In addition to glycoside hydrolases, another α_6/α_6 -barrel enzyme, i.e., *Lactobacillus* maltose phosphorylase, is also an inversion enzyme.

To investigate the two candidates (Asp-88 and Asp-149 for catalytic residues), two mutants (D88N and D149N) having a substitution of Asn for Asp were prepared and assayed (Table 4). CD spectra for wild-type and mutant enzymes showed almost the same profiles (data not shown), indicating that both mutants have no significant conformational change compared to the wild-type enzyme. Specific activity (k_{cat}/K_m) of D88N ($2.9 \times 10^{-6} \mu\text{M}^{-1}\text{s}^{-1}$) and D149N ($9.8 \times 10^{-5} \mu\text{M}^{-1}\text{s}^{-1}$) was, however, significantly lower than that of the wild-type enzyme ($8.1 \times 10^{-2} \mu\text{M}^{-1}\text{s}^{-1}$) for gellan tetrasaccharide (Δ GlcA-Glc-Rha-Glc). Both mutants are almost inactive. The k_{cat} values of D88N (0.00057 s^{-1}) and D149N (0.0059 s^{-1}) were reduced by ~1000-10,000-fold compared to that of the wild-type enzyme (7.3 s^{-1}), while few fluctuations were observed in K_m values among the three enzymes (wild-type, 90 μM ; D88N, 200 μM ; D149N, 60 μM). These kinetics for D88N and D149N suggest that Asp-88 and Asp-149 are essential for catalysis and support the hypothesis postulated by

X-ray crystallography that one acts as the nucleophilile/base and the other as the proton donor/acid.

Intriguingly, the active site of UGL resembles that of hypothetical protein YteR (Fig. 5) more than that of any other structurally similar six-hairpin enzyme. The amino acid sequence identity is very low (less than 10%) between UGL

and YteR, however, and long loops are arranged differently, as discussed above (Fig. 3). Asp-88 in UGL corresponds to Asp-88 in YteR, and in the same way, Asp-149 to Asp-143, His-193 to His-189, Arg-221 to Arg-213, Trp-134 to Trp-141, Trp-219 to Trp-211, and Trp-225 to Trp-217. Trp-42 in UGL is replaced by Tyr-41 in YteR. The side-chains of residues in YteR that correspond to those of His-86 and His-87 in UGL are in opposite directions. Instead of these residues, the space of His-86 and His-87 in UGL is occupied by His-132 and Lys-133 in YteR. The space corresponding to Phe-91 in UGL is also occupied by Met-147 in YteR. The space corresponding to the side-chains of the four residues, His-210, Gln-211, Tyr-338, and His-339, is occupied by the five sequential 331-335 residues of the long loop, L-H11:H12, in YteR. Due to this protruding loop, the putative active site of YteR is smaller than that of UGL. This high structural similarity of putative active sites suggests that YteR binds to a similar UGL substrate (or ligand) or exhibits similar activity.

Table 4. Kinetic parameters for wild-type and mutant UGL.

Enzyme	Substrate	<i>k</i> _{cat} (s ⁻¹)	<i>K</i> _m (μM)	Relative	
				<i>k</i> _{cat} / <i>K</i> _m (μM ⁻¹ s ⁻¹)	<i>k</i> _{cat} / <i>K</i> _m (%)
Wild-type	ΔGlcA-Glc-Rha-Glc	7.3 ± 0.3	90 ± 11	0.081 ± 0.027	100
	ΔGlcA-GlcNAc	5.0 ± 0.3	226 ± 28	0.022 ± 0.011	27
	ΔGlcA-GalNAc	15 ± 0.9	200 ± 26	0.075 ± 0.035	93
D88N	ΔGlcA-Glc-Rha-Glc	0.00057 ± 0.000052	200 ± 38	0.000029 ± 0.000014	0.0036
D149N	ΔGlcA-Glc-Rha-Glc	0.0059 ± 0.00018	60 ± 5.8	0.000098 ± 0.000031	0.12

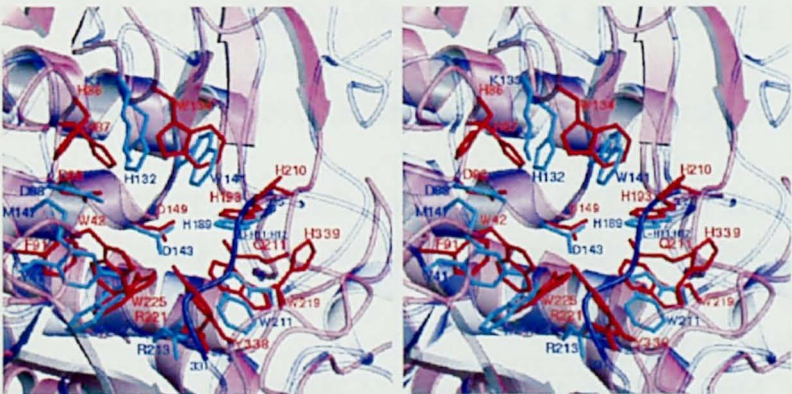


Fig. 5. Structural comparison of the active site arrangement of UGL and hypothetical protein YteR. Main-chains are pink for UGL and blue for YteR. Side-chains are red for UGL and cyan for YteR.

References

- Aleshin, A. E., Feng, P.-H., Honzatko, R. B., and Reilly, P. J. (2003) *J. Mol. Biol.* **327**, 61-73
- Aleshin, A. E., Stoffer, B., Firsov, L. M., Svensson, B., and Honzatko, R. B. (1996) *Biochemistry* **35**, 8319-8328
- Aleshin, A. E., Golubev, A., Firsov, L. M., and Honzatko, R. B. (1992) *J. Biol. Chem.* **267**, 19291-19298
- Alzari, P. M., Souchon, H., and Dominguez, R. (1996) *Structure* **4**, 265-275
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) *Nucl. Acids Res.* **28**, 235-242
- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905-921
- Davies, G., and Henrissat, B. (1995) *Structure* **3**, 853-859
- Egloff, M. P., Uppenberg, J., Haalck, L., and Van Tilbeurgh, H. (2001) *Structure* **9**, 689-697
- Féthière, J., Eggimann, B., and Cygler, M. (1999) *J. Mol. Biol.* **288**, 635-647
- Furey, W., and Swaminathan, S. (1997) *Methods Enzymol.* **277**, 590-620
- Guerin, D. M. A., Lascombe, M.-B., Costabel, M., Souchon, H., Lamzin, V., Beguin, P., and Alzari, P. M. (2002) *J. Mol. Biol.* **316**, 1061-1069
- Hashimoto, W., Kobayashi, E., Nankai, H., Sato, N., Miya, T., Kawai, S., and Murata, K. (1999) *Arch. Biochem. Biophys.* **368**, 367-74
- Hashimoto, W., Maesaka, K., Sato, N., Kimura, S., Yamamoto, K., Kumagai, H., and Murata, K. (1997) *Arch. Biochem. Biophys.* **339**, 17-23
- Hashimoto, W., Nankai, H., Mikami, B., and Murata, K. (2003) *J. Biol. Chem.* **278**, 7663-7673
- Holm, L., and Sander, C. (1993) *J. Mol. Biol.* **233**, 123-138
- Hooft, R. W., Vriend, G., Sander, C., and Abola, E. E. (1996) *Nature* **381**, 272
- Huang, W., Boju, L., Tkalec, L., Su, H., Yang, H. O., Gunay, N. S., Linhardt, R. J., Kim, Y. S., Matte, A., and Cygler, M. (2001) *Biochemistry* **40**, 2359-2372
- Huang, W., Matte, A., Li, Y., Kim, Y. S., Linhardt, R. J., Su, H., and Cygler, M. (1999) *J. Mol. Biol.* **294**, 1257-1269
- Itoh, T., Mikami, B., Maru, I., Ohta, Y., Hashimoto, W., and Murata, K. (2000) *J. Mol. Biol.* **303**, 733-744
- Jedrzejewski, M. J., Mello, L. V., de Groot, B. L., and Li, S. (2002) *J. Biol. Chem.* **277**, 28287-28297
- Juy, M., Amit, A. G., Alzari, P. M., Poljak, R. J., Claeysens, M., Bguin, P., and Aubert, J. (1992) *Nature* **357**, 89-91
- Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24**, 946-950
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26**, 283-291
- Li, S., and Jedrzejewski, M. J. (2001) *J. Biol. Chem.* **276**, 41407-41416

- Li, S., Kelly, S. J., Lamani, E., Ferraroni, M., and Jedrzejas, M. J. (2000) *EMBO J.* **19**, 1228-1240
- Linker, A., Meyer, K., and Weissmann, B. (1955) *J. Biol. Chem.* **213**, 237-248
- Lunin, V. V., Li, Y., Linhardt, R. J., Miyazono, H., Kyogashima, M., Kaneko, T., Bell, A. W., and Cygler, M. (2004) *J. Mol. Biol.* **337**, 367-386
- Luzzati, V. (1952) *Acta Crystallogr.* **5**, 802-810
- Maruyama, Y., Hashimoto, W., Mikami, B., and Murata, K. (2005) *J. Mol. Biol.* **350**, 974-986
- McCarter, J. D., and Withers, S. G. (1994) *Curr. Opin. Struct. Biol.* **4**, 885-892
- Mello, L. V., De Groot, B. L., Li, S., and Jedrzejas, M. J. (2002) *J. Biol. Chem.* **277**, 36678-36688
- Merrit, E. A., and Murphy, M. E. P. (1994) *Acta Crystallogr. D Biol. Crystallogr.* **50**, 869-873
- Mori, S., Akao, S., Miyake, O., Nankai, H., Hashimoto, W., Mikami, B., and Murata, K. (2003a) *Acta Crystallogr. D Biol. Crystallogr.* **59**, 946-949
- Mori, S., Akao, S., Nankai, H., Hashimoto, W., Mikami, B., and Murata, K. (2003b) *Protein Expr. Purif.* **29**, 77-84
- Murzin, A. Z., Brenner, S. E., Hubbard, T. J., and Chothia, C. (1995) *J. Mol. Biol.* **247**, 536-540
- Nicholls, A., Sharp, K., and Honig, B. (1991) *Proteins Struct. Funct. Genet.* **11**, 281-296
- Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276**, 307-326
- Parsieglä, G., Reverbel-Leroy, C., Tardif, C., Belaich, J. P., Driguez, H., and Haser, R. (2000) *Biochemistry* **39**, 11238-11246
- Ponnuraj, K., and Jedrzejas, M. J. (2000) *J. Mol. Biol.* **299**, 885-895
- Sakon, J., Irwin, D., Wilson, D. B., and Karplus, P. A. (1997) *Nature Struct. Biol.* **4**, 810-818
- Shaya, D., Tocilj, A., Li, Y., Myette, J., Venkataraman, G., Sasisekharan, R., and Cygler, M. (2006) *J. Biol. Chem.* **281**, 15525-15535
- Ševčík, J., Solovíková, A., Hostinová, E., Gašperík, J., Wilson, K. S., and Dauter, Z. (1998) *Acta Crystallogr. D Biol. Crystallogr.* **54**, 854-866
- Smith, N. L., Taylor, E. J., Lindsay, A. M., Charnock, S. J., Turkenburg, J. P., Dodson, E. J., Davies, G. J., and Black, G. W. (2005) *Proc. Natl. Acad. Sci. USA* **102**, 17652-17657
- Van Petegem, F., Collins, T., Meuwis, M. A., Gerday, C., Feller, G., and Van Beeumen, J. (2003) *J. Biol. Chem.* **278**, 7531-7539
- Wang, B. C. (1985) *Methods Enzymol.* **115**, 90-112

CHAPTER II

A NOVEL GLYCOSIDE HYDROLASE FAMILY 105

Pectin, one of the most complex polysaccharides in plant cell walls, contains both linear (polygalacturonate) and ramified (rhamnogalacturonan, RG) regions (Albersheim *et al.*, 1958). Two types of ramified region have been identified, i.e., RG-I and RG-II (Schols *et al.*, 1995). RG-I has a backbone of a repeating disaccharide unit that consists of galacturonic acid (GalA) and rhamnose (Rha) residues and is fully depolymerized by a range of enzyme activities (Fig. 1). Long chains of neutral polymers (arabinans and galactans) are attached as side-chains to about 50% Rha present in RG-I (McNeil *et al.*, 1984). RG-II is a very complex polysaccharide containing some 30 different monosaccharides, including GalA, Rha, and several uncommon saccharides (O'Neill *et al.*, 1996). GalA present in RG-I and RG-II is often acetylated at the O2 or O3 positions.

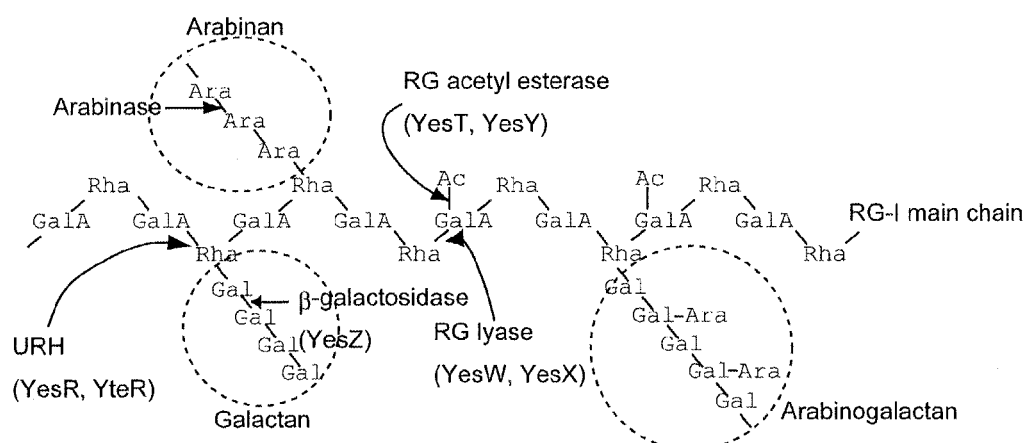


Fig. 1. Postulated RG-I-degrading pathway in *B. subtilis* strain 168. Cleavage sites of RG-I depolymerization enzymes (URH, RG lyase, RG acetyl esterase, arabinase, and galactosidase) are indicated by arrows. RG lyases (YesW and YesX) produce unsaturated RG from the RG-I main-chain. URHs (YesR and YteR) hydrolyze unsaturated RG. RG acetyl esterases (YesT and YesY) possibly release the acetyl (Ac) group from the acetylated GalA. The RG side-chains such as arabinan and galactan are degraded by arabinase and β -galactosidases.

Many pathogenic plant bacteria produce enzymes responsible for the degradation of plant cell walls containing RG (Salmond, 1994). Among such bacteria, the enterobacterium *Erwinia chrysanthemi* has fully been characterized for plant cell wall degradation. The bacterium causes soft rot diseases in different plants (Bauer *et al.*, 1994) and produces several types of pectin (or

RG)-related active enzymes such as pectin methylesterases, pectin (or RG) acetylerases, exo- and endopectate (or RG) lyases, exopolygalacturonases, and exooligogalacturonate lyase. (Hugouvieux-Cotte-Pattat *et al.*, 1996; Shevchik *et al.*, 1999). Endopectate (or RG) lyases are the main virulence factors in bacterial pathogens that cause maceration. Through β -elimination, pectate (or RG) lyases recognize GalA present in polysaccharides and cause the release of saccharides with unsaturated galacturonic acid (Δ GalA, 4-deoxy- β -L-threo-hex-4-enopyranosyl uronic acid) that have a C=C double bond at the nonreducing terminus (Barras *et al.*, 1994). Exooligogalacturonate lyase acts on unsaturated oligogalacturonate (Δ GalA-(GalA)_n) produced from polygalacturonate by the action of pectate lyases and releases Δ GalA via β -elimination (Shevchik *et al.*, 1999). No reports have been made, however, on “unsaturated galacturonyl hydrolase” (URH) catalyzing the cleavage of the glycoside bond between Δ GalA and sugars through hydrolysis.

Typical microbial pathogens invade mammalian tissues by degrading extracellular matrix “glycosaminoglycans” via the action of polysaccharide lyases and hydrolases similar to plant pathogens that invade plant cell walls (Sawitzky, 1996; Boulnois, 1992). YteR, a hypothetical protein with unknown functions, is derived from *Bacillus subtilis* strain 168, and its overall structure, with the exception of some loops, significantly resembles that of UGL (Figs. 3 and 5, Chapter I), although it exhibits little amino acid sequence identity with UGL (Fig. 2). A deep pocket shared in YteR and UGL seems to be an active site. In this pocket, possible residues involved in substrate recognition and catalysis are conserved in YteR and UGL (Fig. 5, Chapter I). YteR family proteins are highly conserved in bacteria and fungi, and its crystal structure was determined in a structural genomics project at the Midwest Center (Zhang *et al.*, 2005). In *B. subtilis* strain 168, another gene coding for YesR, which is homologous to YteR, is also present.

This chapter deals with the identification of YteR and its homologous protein YesR as URH and the characterization of these proteins responsible for the degradation of RG-I in *B. subtilis* strain 168.

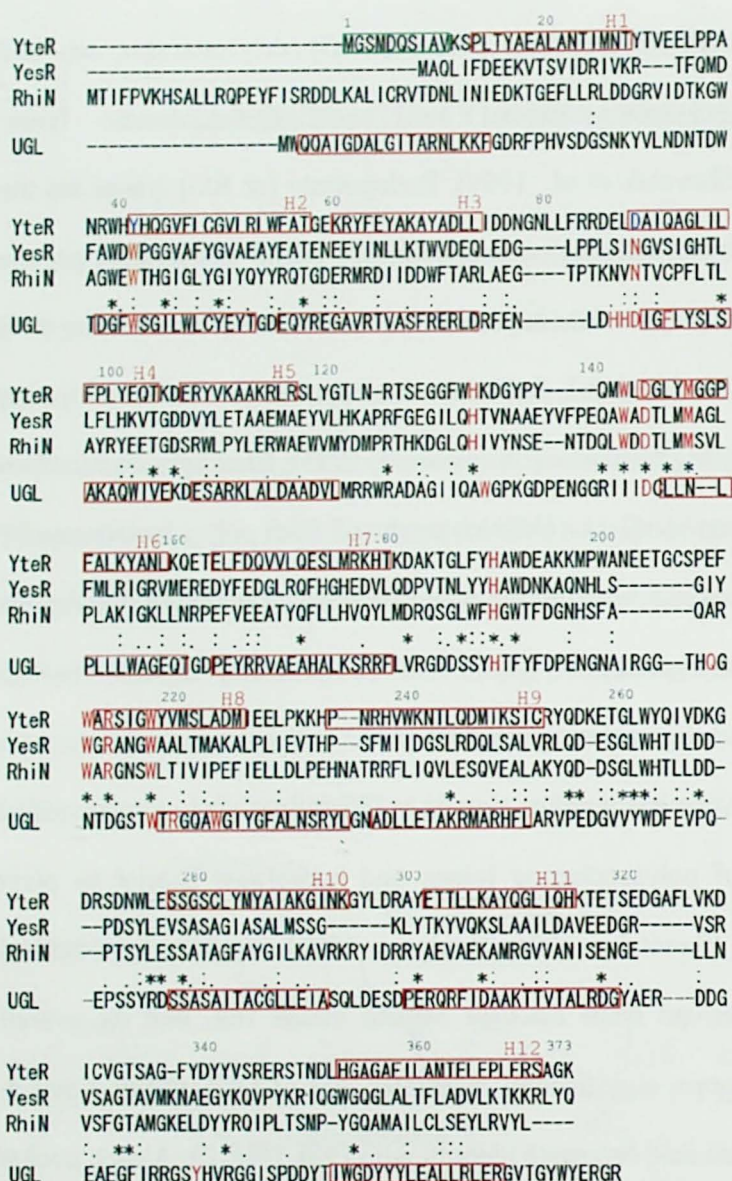


Fig. 2. Amino acid sequence alignment of YteR and other homologous proteins using ClustalW program analysis (Thompson *et al.*, 1994). YteR, YteR protein of *B. subtilis* strain 168 (Accession No., CAB14990); YesR, YesR protein of *B. subtilis* strain 168 (Accession No., CAB12519); RhiN, RhiN protein of *E. chrysanthemi* (Accession No., CAC83616); UGL, unsaturated glucuronyl hydrolase of *Bacillus* sp. GL1 (Accession No., BAA84216). Identical amino acid residues among the three proteins are indicated by asterisks. The green box indicates invisible residues obtained by X-ray crystallographic analysis. α -helices are indicated by red boxes. Residues marked in red or blue are located at the active pocket.

Materials and Methods

Molecular cloning of YteR and YesR genes

A genomic DNA of *B. subtilis* strain 168 was isolated based on the method used generally (Sambrook *et al.*, 1989). To subclone YteR and YesR genes into expression vector pET21d (Novagen, WI, USA), polymerase chain reaction (PCR) was done using KOD plus polymerase (Toyobo, Tokyo, Japan), genomic DNA as a template, and two synthetic oligonucleotides with *Nco*I and *Xho*I restriction sites added to their termini as forward and reverse primers. The nucleotide

sequences of the primers used in PCR were 5'-GGATCCCATGGGATCAATGGATCAATCAATCGCTG-3' (forward) and 5'-CCGCTCGAGTCATTTCCCGCTGACCTAAAAAGAG-3' (reverse) for the YteR gene and 5'-GGATCCCATGGCACAGCTTATCTTTGATGAAGAAAAG-3' (forward) and 5'-CCGCTCGAGTTACTGATACAAGCGTTTTTTCGTTTTTC-3' (reverse) for the YesR gene; the residues in bold letters indicate recognition sites for *Nco*I or *Xho*I. Amplified genes were digested with *Nco*I and *Xho*I, then ligated with the *Nco*I- and *Xho*I-double-digested expression vector. Resultant plasmids containing YteR and YesR genes were designated pET21d-YteR and pET21d-YesR. The accuracy of the two genes was confirmed by DNA sequencing. Phylogenetic analysis was done using the BLAST (Altshul *et al.*, 1990) and ClustalW (Thompson *et al.*, 1994) programs on the DDBJ server (Shizuoka, Japan).

Purification of YteR and YesR

E. coli BL21(DE3) was used as a host cell for the overexpression of YteR and YesR. To purify enzymes expressed in *E. coli* that harbored the plasmid (pET21d-YteR or pET21d-YesR), cells were aerobically precultured at 30°C in 1.5 L of Luria-Bertani medium. When the turbidity at 600 nm reached 0.5, the culture was supplemented with 1 mM isopropyl- β -D-thiogalactopyranoside, followed by further incubation at 18°C for 15 h. Cultured cells were collected by centrifugation at 6000 g and 4°C for 5 min and washed with a buffer containing 20 mM Tris-HCl, pH 7.5 (Tris), 1 mM phenylmethylsulfonyl fluoride, and 0.1 μ M pepstatin A; cells were then resuspended in the same buffer and ultrasonically disrupted (Insonator Model 201M, Kubota, Tokyo, Japan) at 0°C and 9 kHz for 20 min. The clear solution obtained after centrifugation at 15,000 g and 4°C for 20 min was used as the cell extract. The cell extract was applied to a HiLoad 16/10 Q Sepharose HP column (Amersham Biosciences Corp., NJ, USA) that was equilibrated with 20 mM Tris. Enzymes were eluted with a linear gradient of sodium chloride (0-1.0 M) in 20 mM Tris (480 ml). YteR and YesR were eluted with approximately 0.5 M sodium chloride in 20 mM Tris. Fractions containing

enzymes were dialyzed overnight against 20 mM Tris at 4°C and applied to a Mono Q 10/10 HP column (Amersham Biosciences Corp., NJ, USA) equilibrated with 20 mM Tris. Enzymes were eluted with a linear gradient of sodium chloride (0-0.7 M) in 20 mM Tris (180 ml). YteR and YesR were eluted with approximately 0.5 M sodium chloride in 20 mM Tris. Fractions containing enzymes were combined, saturated with ammonium sulfate (30%), then applied to a Butyl-Toyopearl 650 M column (1.6 × 8 cm; Tosoh, Tokyo, Japan) that was equilibrated with 20 mM Tris saturated with ammonium sulfate (30%). Enzymes were eluted with a linear gradient of ammonium sulfate (30%-0%) in 20 mM Tris (300 ml). YteR and YesR were eluted with 20 mM Tris saturated with ammonium sulfate (approximately 20%-25%). Enzyme solutions were dialyzed overnight against 20 mM Tris at 4°C and used as purified enzyme sources. Enzyme purity was assessed by SDS-PAGE followed by Coomassie Brilliant Blue R-250 staining. N-terminal amino acid sequences of these proteins were determined by Edman degradation using Procise 492 protein sequencing (Applied Biosystems, CA, USA).

Preparation of unsaturated RG disaccharide

Unsaturated RG disaccharide as a substrate for URH was prepared using potato RG-I (Megazyme International Ireland Ltd., Wicklow, Ireland) and RG lyases (YesW and YesX from *B. subtilis* strain 168) as described below. RG-I was saponified with 50 mM sodium hydroxide at 100°C for 15 min. Saponified and neutralized RG-I was hydrolyzed with 50 mM hydrochloric acid at 80°C for 72 h. The RG-I main-chain was isolated by using a Q-Sepharose Fast Flow anion-exchange column (1.6 × 10 cm; Amersham Biosciences Corp., NJ, USA) with a linear gradient of ammonium bicarbonate (0.01-1.0 M) in distilled water (400 ml) and was depolymerized by RG lyases (YesW and YesX). Unsaturated RG saccharides with varying oligomeric degrees were separated from the polysaccharide by using a Bio-Gel P2 gel permeation column (1.6 × 120 cm; Bio-Rad Laboratories, CA, USA) that was equilibrated with distilled water. The unsaturated RG disaccharide (Δ GalA-Rha) fraction was purified further by an aminopropyl normal-phase column

(Cosmosil 5NH₂-MS, Nacalai Tesque, Kyoto, Japan) that was equilibrated with 70% (v/v) acetonitrile. The purity of saccharide samples was assessed by thin layer chromatography (TLC) as described below.

TLC

RG oligomers and degradation products of Δ GalA-Rha obtained by the action of URH were analyzed on a TLC plate (E. Merck, Darmstadt, Germany) with a solvent system of 1-butanol:acetic acid:water (3:2:2, v/v). Products were visualized by heating the plate at 110°C for 5 min after spraying with 10% (v/v) sulfuric acid in ethanol.

Detection of α -keto acid

The reaction with thiobarbituric acid (TBA) for detection of α -keto acid was as follows (Weissback & Hurwitz, 1959): Products (0.1 ml) from Δ GalA-Rha by the action of URH were added to 0.1 ml of 2% sodium acetate and 0.5 N HCl. After 0.4 ml of 0.3% TBA was added, the mixture was heated at 100°C for 10 min. The TBA-reactive product was determined by measuring the absorbance at 548 nm and was also separated on a TLC plate. The TBA reaction (0.1 ml) was also done with products from Δ GlcA-GalNAc by the action of UGL, and the resultant TBA-reactive product was used as a TLC standard. The reaction with semicarbazide was done as follows (MacGee & Doudoroff, 1954): Products (0.1 ml) from Δ GalA-Rha by the action of URH were added to 0.1 ml of 1% semicarbazide hydrochloride and 1.5% sodium acetate. The reaction mixture was incubated at 30°C for 15 min, diluted to 0.5 ml with water, and subjected to measurement of the absorbance at 250 nm.

Enzyme assay

URH reactions for YteR and YesR were conducted at 30°C as follows: The reaction mixture (0.5 ml) consisted of 50 mM sodium acetate (pH 4.0) for YteR or sodium phosphate (pH 6.0) for YesR, the substrate, and corresponding enzyme. Enzyme activity was measured by monitoring the decrease in absorbance at 235 nm; this corresponded to the loss of the C=C double

bond of the substrate (molar extinction coefficient $\epsilon_{235} = 5,200 \text{ (M}^{-1}\text{cm}^{-1})$) (McKie *et al.*, 2001) because the pyranose ring of the released ΔGalA readily opens, enabling its conversion to α -keto acid due to the loss of the double bond (Hashimoto *et al.*, 1999; Linker *et al.*, 1955). The protein concentration was determined by UV spectrophotometry using the theoretical molar extinction coefficient $\epsilon_{280} = 94,590 \text{ (M}^{-1}\text{cm}^{-1})$ for YteR or $71,690 \text{ (M}^{-1}\text{cm}^{-1})$ for YesR. Unsaturated hyaluronate disaccharide ($\Delta\text{GlcA-GlcNAc}$) and unsaturated chondroitin disaccharide ($\Delta\text{GlcA-GalNAc}$) produced by the action of polysaccharide lyases were obtained from Seikagaku Corporation (Tokyo, Japan). Unsaturated oligogalacturonate ($\Delta\text{GalA-(GalA)}_n$) was prepared as described previously (Hashimoto *et al.*, 1999). Parameters k_{cat} and K_{m} were determined by nonlinear fitting to the Michaelis-Menten equation. Optimal pH, temperature, and thermal stability were evaluated based on $k_{\text{cat}}/K_{\text{m}}$. When the substrate concentration $[\text{S}]$ was significantly lower than K_{m} ($[\text{S}] \ll K_{\text{m}}$), reactions followed the conditions of first-order reactions. First-order rate constants were converted to second-order rate constants ($k_{\text{cat}}/K_{\text{m}}$). The optimal pH of enzyme reactions was determined using 50 mM sodium acetate (pH 3.5-6.0) for YteR or 50 mM sodium phosphate (pH 4.8-8.2) for YesR. The thermal stability of enzymes was determined after preincubating purified enzymes at temperatures ranging from 30 to 60°C for 10 min, then measuring the remaining enzymatic activity as described. To determine the optimal temperature, reactions were done at temperatures ranging from 30 to 65°C.

Crystallization and structure determination

YteR was crystallized by hanging drop vapor diffusion described by Zhang *et al.* (2005). The YteR crystal was soaked in a solution containing 400 mM $\Delta\text{GlcA-GalNAc}$ for 15 min at 20°C. The solution was prepared in 1.2 M citrate and 0.1 M Tris. The crystal was placed in a cold nitrogen gas stream at -173°C. X-ray diffraction images were collected using a Jupiter210 area detector (Rigaku, Tokyo, Japan) with synchrotron radiation at a wavelength of 0.80 Å at the BL-38B1 station of SPring-8 (Hyogo, Japan). Images were processed to a resolution of 1.7 Å using DENZO

and SCALEPACK software (Table 1) (Otwinowski & Minor, 1997). The YteR crystal structure in complex with Δ GlcA-GalNAc was determined by molecular replacement using the CNS ver. 1.1 program (Brünger *et al.*, 1998) with the previously reported YteR structure as the reference model. Coordinates of UGL (1VD5), UGL mutant D88N/ Δ GlcA-GalNAc complex (2AHG), and YteR (1NC5) were obtained from the RCSB Protein Data Bank (<http://pd-beta.rcsb.org>). Several rounds of refinement followed by manual model building using the TURBO-FRODO

Table 1. Data collection and refinement statistics for a YteR/ Δ GlcA-GalNAc crystal

Data collection	
Resolution limit (last shell) ^a (Å)	50.0-1.70 (1.76-1.70)
Measured reflections	504184
Unique reflections (last shell)	47953 (4772)
Redundancy (last shell)	10.5 (10.1)
Completeness ($ I > \sigma(I)$) (last shell) (%)	98.1 (96.3)
R_{merge} (%) ^b	5.4 (35.7)
Refinement	
Final model	363 residues (11-373), 675 water, and Δ GlcA-GalNAc molecules
Resolution limit (last shell) (Å)	10.0-1.70 (1.76-1.70)
Used reflections (last shell)	45283 (4458)
Completeness ($F > \sigma(F)$) (last shell) (%)	91.0 (93.1)
Average B -factor (Å ²)	24.2
Protein	19.1
Water	45.1
Δ GlcA	27.4
GalNAc	81.8
R -factor (last shell) (%) ^c	16.2 (28.5)
R_{free} (%) ^d	21.9
rms deviations	
Bond (Å)	0.008
Angle (Å)	0.026
Ramachandran plot (%)	
Most favored	95.7
Disallowed	0.0

^a Data of highest resolution shells is provided in parentheses.

^b $R_{\text{merge}} = \sum |I_i - \langle I \rangle| / \sum I_i \times 100$, where I_i is the intensity of individual reflection and $\langle I \rangle$ the mean intensity of all reflections.

^c R -factor = $\sum |F_o - F_c| / \sum |F_o| \times 100$, where F_o is the observed structure factor and F_c the calculated structure factor.

^d R_{free} was calculated from randomly chosen 5% reflections.

program (AFMB-CNRS, Marseille, France) were done to improve the model by increasing data resolution to 1.70 Å using the SHELXL program (Sheldrick & Schneider, 1997). The final R -factor was 16.2% for 45,283 data points with a resolution of 10.0-1.70 Å (91.0% completeness). The R -free value calculated for the randomly separated 5% data was 21.9 (Table 1). Structural alignment was done by superimposing using a fitting program in TURBO-FRODO. Ribbon plots were prepared using the MOLSCRIPT (Kraulis, 1991) and Raster3D (Merrit & Murphy, 1994) programs.

Results

Genetic analysis of YteR

Although little amino acid sequence identity exists between YteR and UGL (<10%), their three-dimensional structures are very similar (root mean square distance = 2.80 Å for the superimpositioning of 323 Cα; Figs. 3 and 5, Chapter I). Zhang *et al.* have shown that 71 proteins with unknown functions belong to the YteR family and are widely conserved in bacteria and fungi (2005). A large number of glycoside hydrolases containing UGL have been analyzed so far, and based on their primary structure, most can be classified into more than 100 families in the CAZy database (Coutinho & Henrissat, 1999). The amino acid sequence of YteR (or YteR family), however, shows no homology with these glycosidases in the CAZy database.

In the YteR family, BLAST and ClustalW programs showed that the amino acid sequence of YteR from *B. subtilis* strain 168 shows homology those of YesR from *B. subtilis* strain 168 (26% identity in 306 amino acid residues overlap; Accession No., CAB12519) and *E. chrysanthemi* RhiN (26% identity in 332 amino acid residues overlap; Accession No., CAC83616) (Fig. 2). Information on the functions of YteR family proteins is limited, with the exception of RhiN, which has been shown to reduce the absorbance of RG derivatives at 235 nm, probably of unsaturated RG, after RG is treated with RG lyase (Hugouvieux-Cotte-Pattat, 2004). Although detailed kinetic studies on RhiN are insufficient and results are extremely preliminary, RhiN appears to be involved in the degradation of unsaturated RG during RG metabolism. It has been predicted that hypothetical proteins coded by a genetic cluster including the YesR gene are related to RG metabolism in *B. subtilis* strain 168, e.g., the RG lyases, namely, YesW and YesX (polysaccharide lyase family 11) (Pages *et al.*, 2003) and the RG acetyl esterases, namely, YesT and YesY (carbohydrate esterase family 12) (Molgaard *et al.*, 2000). They appear to form a single cluster for RG-I depolymerization. Based on this information on the *Erwinia* RhiN and the *Bacillus* genetic cluster, the identification of YteR family proteins as enzymes responsible for RG metabolism was carried out.

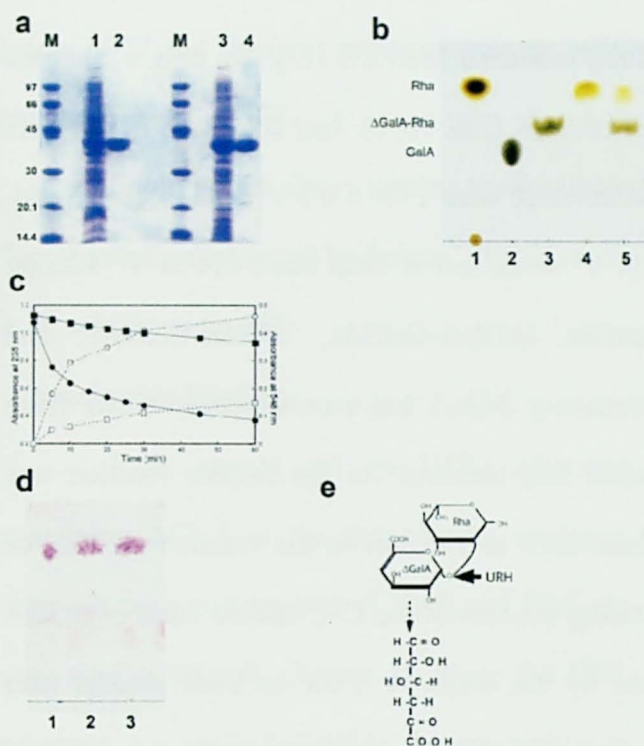


Fig. 3. Functional expression of YteR and YesR. (a) SDS-PAGE profile of YteR and YesR. Purified YteR and YesR were subjected to SDS-PAGE (11% gel). Protein bands were stained with Coomassie Brilliant Blue R-250. Lane M, molecular weight standards; lane 1, cell extract of *E. coli* harboring pET21d-YteR; lane 2, purified YteR; lane 3, cell extract of *E. coli* harboring pET21d-YesR; lane 4, purified YesR (b) Reaction products incubated with ΔGalA-Rha and the corresponding enzyme were subjected to TLC analysis. Lane 1, 25 μg of standard Rha; lane 2, 25 μg of standard GalA; lane 3, 50 μg substrate ΔGalA-Rha; lane 4, reaction products after incubation at 30°C for 60 min with YesR (1 μg) and ΔGalA-Rha (50 μg); lane 5, reaction products after incubation at 30°C for 60 min with YteR (1 μg) and ΔGalA-Rha (50 μg). (c) Monitoring of YteR (1 μg) and YesR (1 μg) reactions with ΔGalA-Rha (50 μg) at 30°C. Decrease in absorbance at 235 nm during the reaction (YteR (■) and YesR (●)). Increase in reaction product (α-keto acid) detected by TBA in absorbance at 548 nm during the reaction ((YteR (□) and YesR (○)). (d) Reaction products (TBA complex) were subjected to TLC analysis. Lane 1, 0.1 μg of standard ΔGlcA-TBA complex (UGL reaction product-TBA complex); lane 2, 0.1 μg of YesR reaction product-TBA complex; lane 3, 0.1 μg YesR reaction product-TBA complex. (e) unsaturated RG disaccharide (ΔGalA-Rha).

Expression and purification of YteR and YesR

The *E. coli* BL21(DE3) transformant that expresses the plasmid (pET21d-YteR or pET21d-YesR) under control of the T7 promoter and terminator produced a large amount (~20%) of YteR or YesR protein within bacterial cells (Fig. 3a). YteR and YesR purified to homogeneity from *E. coli* cells showed monomeric forms with molecular masses of 43 (YteR) and 39 (YesR) kDa as determined by SDS-PAGE (Fig. 3a) and gel permeation chromatography (data not shown); this observation corresponds to molecular masses calculated from their predicted open reading frames. N-terminal amino acid sequences of the proteins were determined to be NH₂-GSMDQ (YteR) and NH₂-AQLIF (YesR), which also correspond to their predicted sequences (Fig. 2), although their first amino acid residues (Met-1) were excised by *E. coli* amino peptidase.

Characterization of YteR and YesR

To clarify their functions, YteR and YesR were subjected to enzyme assay in a mixture containing various unsaturated disaccharide substrates. YteR and YesR were found to specifically

degrade unsaturated RG disaccharide (Δ GalA-Rha) produced from RG (Fig. 3b, c, d, e), but the activity of YteR was found to be lower than that of YesR (Fig. 3b, c). k_{cat} and K_m of YteR were $0.28 \pm 0.011 \text{ s}^{-1}$ and $100 \pm 14 \text{ }\mu\text{M}$, while those of YesR were $13.9 \pm 0.7 \text{ s}^{-1}$ and $719 \pm 75 \text{ }\mu\text{M}$. Specific activity (k_{cat}/K_m) of YteR was $0.0028 \text{ M}^{-1}\text{s}^{-1}$, while that of YesR was $0.014 \text{ M}^{-1}\text{s}^{-1}$. Neither enzyme acts on other unsaturated saccharides (Δ GlcA-GalNAc, Δ GlcA-GlcNAc, and Δ GalA-(GalA)_n), which are UGL substrates containing Δ GlcA and exooligogalacturonate lyase substrate containing Δ GalA. The resultant product Rha obtained via the enzyme reaction was detected by using TLC analysis (Fig. 3b). Absorbance at 235 nm in the reaction mixture of Δ GalA-Rha with YteR or YesR decreased, indicating the loss of the C=C double bond of Δ GalA (Fig. 3c). The product derived from Δ GalA-Rha by the action of YteR or YesR reacted with semicarbazide and the resultant reaction product showed significant absorbance at 250 nm, which is specific to α -keto acid. α -Keto acid was also detected as a TBA-reactive compound showing an absorbance at 548 nm during the reaction (Fig. 3c). R_f of the enzyme (YteR or YesR) reaction product complexed with TBA was identical to that of the UGL reaction product (Δ GlcA) complexed with TBA (Fig. 3d). Δ GalA is identical to Δ GlcA due to the absence of the hydroxyl group at the C4 position and spontaneously converted to the chain form, because the ring of Δ GlcA (Δ GalA) is unstable due to keto-enole equilibrium (Fig. 3e) (Hashimoto *et al.*, 1999; Linker *et al.*, 1955). These analyses indicated that products of the action of YteR or YesR were Δ GalA and Rha, meaning that both YteR and YesR are shown to catalyze the hydrolytic reaction of unsaturated disaccharide.

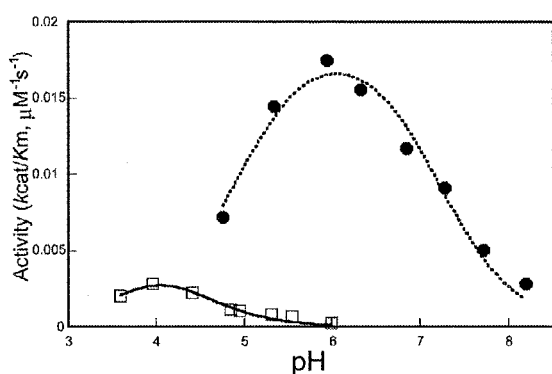


Fig. 4. pH profiles of YteR (□) and YesR (●)

The bell-shaped pH profiles of YteR and YesR are shown in Fig. 4. Similar to UGL, YesR showed maximum activity at pH 6. pH 4 was the optimal pH for the YteR reaction; this was lower than that required for the UGL and YesR reaction. YteR and YesR were stable

below 50°C and showed maximum activity at 50°C. Temperature profiles of YteR and YesR are similar to those of UGL.

Based on the above results, YteR and YesR were identified to constitute a novel enzyme, “unsaturated galacturonyl hydrolase”, that is responsible for the degradation of the RG-I main-chain in *B. subtilis* strain 168.

Saccharide binding

YteR cannot act on UGL substrates containing ΔGlcA at the nonreducing terminus, although the active pockets of YteR and UGL are very similar (Fig. 5, Chapter I). Since the UGL substrate, e.g., ΔGlcA-GalNAc, is very similar to the YteR substrate (ΔGalA-Rha), a crystal of YteR in complex with a pseudosubstrate

(ΔGlcA-GalNAc) (Fig. 5a) was prepared according to the crystallization condition described by Zhang *et al.* (2005). Although YteR has the binding ability of ΔGlcA-GalNAc at the gate of the deep pocket, the arrangement of the saccharide in YteR/ΔGlcA-GalNAc differs from that of the UGL inactive mutant D88N/substrate complex (Fig. 5b).

The apo and holo forms of YteR have essentially identical structures, with no conformational change observed in the binding of ΔGlcA-GalNAc. The root mean square distance was 0.06 Å for the superimposition of all residues (363 Cα atoms). In the deep pocket, they also share almost similar structures with their side-chains, while ΔGlcA-GalNAc in its holo form was modeled in the place of some waters of apo form. ΔGlcA interacts with YteR residues—His-40, Tyr-41, His-42, Lys-133, Tyr-136, Trp-141, Gly-332, Thr-333, and Ser-334 (Fig. 5a and Table 2). In particular, hydrogen bonds play an essential role in the binding of ΔGlcA. The positively charged side-chains of His-40 and His-42 face the carboxyl groups of ΔGlcA. The two O atoms (O6A and

Table 2. Interaction between YteR and ΔGlcA

Hydrogen bonds (< 3.3 Å)			
ΔGlcA atom	Paired atom in YteR		Distance (Å)
O2	Lys-133	Nζ	3.1
O2	Tyr-136	Oη	3.2
O3	Ser-334	Oγ	2.6
O6A	Tyr-41	Oη	3.2
O6B	His-40	Ne2	3.0
van der Waals contact (C-C distance < 4.5 Å)			
ΔGlcA atom	Paired residue in YteR		
C2	Tyr-136, Trp-141		
C3	Trp-141, Thr-333, Ser-334		
C4	Gly-332, Thr-333		
C6	His-40, Tyr-41, His-42, Gly-332		

O6B) of the carboxyl group form hydrogen bonds directly with residues. Direct hydrogen bonds are responsible for the association between O2 and O3 atoms and Lys-133, Tyr-136, and Ser-334. The hydrophobic side-chains of Tyr-41, Tyr-136, and Trp-141 partially interact with the Δ GlcA pyranose ring via hydrophobic interactions. GalNAc does not directly interact with the residues (Fig. 5a). Although some hydrogen bonds with water molecules are present, GalNAc is highly flexible and clings to Δ GlcA. This is supported by the fact that the average *B*-factor of Δ GlcA is significantly lower than that of GalNAc (Table 1).

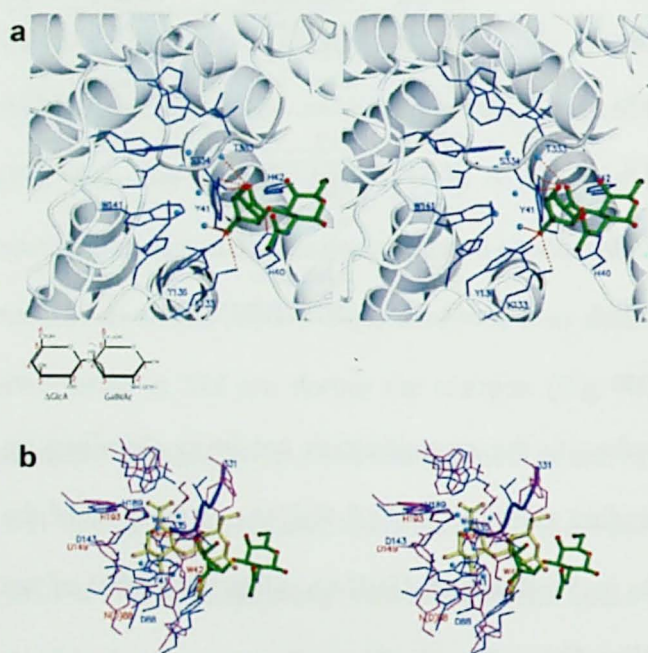


Fig. 5. Crystal structure of YteR complexed with Δ GlcA-GalNAc. (a) Structure of YteR bound with Δ GlcA-GalNAc at the active pocket. The active pocket of α_6/α_6 -barrel is shown in gray. Labeled residues represented by blue bond models interact with Δ GlcA-GalNAc. Δ GlcA-GalNAc is denoted by a green bond model (oxygen atom, red; carbon atom, black; nitrogen atom, blue). Water molecules are represented as cyan balls. Hydrogen bonds are shown as red lines. The numbers of carbon atoms of Δ GlcA-GalNAc is also indicated in a model. (b) Comparison of YteR/ Δ GlcA-GalNAc and UGL (D88N)/ Δ GlcA-GalNAc complex (2AHG) structures at the active site. Amino acid residues are blue (YteR) and magenta (UGL). The blue ribbon model represents the protruding loop of YteR (amino acid residues, 331 through 335). Δ GlcA-GalNAc structures are denoted by green (YteR complex) and yellow (UGL complex) bond models (oxygen atom, red; carbon atom, black; nitrogen atom, blue). Water molecules of YteR/ Δ GlcA-GalNAc (cyan) and UGL (D88N)/ Δ GlcA-GalNAc (red) are represented as ball models.

Discussion

Novel glycoside hydrolase

YteR and YesR were first identified as a novel glycoside hydrolase URH. Since YteR family proteins show little homology with glycoside hydrolases analyzed to date, it is proposed that YteR and YesR constitute a novel glycoside hydrolase family. Henrissat *et al.* recognized this and assigned YteR and YesR to glycoside hydrolase family 105 in the CAZy database (Henrissat, B., Coutinho, P., and Deleury, E.: <http://afmb.cnrs-mrs.fr/CAZY/>).

Active pocket

In the active pocket, residues are highly conserved between YteR and UGL (Fig. 5, Chapter I). The space corresponding to the side-chains of the four residues, His-210, Gln-211, Tyr-338, and His-339 in UGL is occupied by five sequential 331–335 residues of the long loop in YteR (Fig. 5, Chapter I). The five sequential residues, 331 to 335, in YteR reduce the width of the gate of the active pocket (Fig. 5a, b). Due to the protruding loop containing residues from 331 to 335, the +1 subsite of YteR is narrower than that of UGL. Based on the crystal structures of YteR complexed with Δ GlcA-GalNAc and UGL complexed with Δ GlcA-GalNAc, the saccharide binding-ability and active site of the enzyme was clarified (Fig. 5b). The structures provide information on -1 to +2 subsites, although Δ GlcA in YteR/ Δ GlcA-GalNAc occupies the +1 subsite, and not the -1 subsite. The ring of GalNAc in the UGL complex cannot be superimposed on that of Δ GlcA in the YteR complex by rotating along the glycoside bond. The size of GalNAc containing the large functional group, i.e., the *N*-acetyl group, is larger than that of Rha. GalNAc would be sterically hindered by the +1 subsite.

RG-I depolymerization in *B. subtilis* strain 168

Based on phylogenetic analysis and enzyme characterization, the RG-I depolymerization pathway in *B. subtilis* strain 168 is postulated in Fig. 1. The RG-I main-chain is deacetylated by RG acetylsterase (YesT and YesY) (Molgaard *et al.*, 2000) and degraded to unsaturated saccharide by the action of RG lyase (YesW and YesX) (Pages *et al.*, 2003). The Δ GalA residue is released from unsaturated RG by the action of URH (YesR and YteR) that is present for the complete depolymerization of the RG-I main-chain. The RG-I side-chain (galactan, arabinan, and arabinogalactan) is depolymerized by β -galactosidase (YesZ, glycoside hydrolase family 42) (De Vries *et al.*, 2002) and arabinan-endo 1,5- α -L-arabinase (Sakamoto *et al.*, 1997). With the exception of the arabinase and YteR genes, the genes of predicted enzymes, which are responsible for RG-I metabolism, are located closely in the genome. They appear to form a single cluster for RG-I

depolymerization. *E. chrysanthemi* causes soft rot diseases in plants by enzymatically degrading pectin in plant cell walls (Hugouvieux-Cotte-Pattat *et al.*, 1996). RhiE (RG lyase) and RhiN (predicted to be URH) are highly induced during plant infection caused by *E. chrysanthemi* and are believed to participate in RG-I metabolism (Hugouvieux-Cotte-Pattat, 2004). These enzymes of *B. subtilis* strain 168 are thus probably regulated and induced during plant infection and RG-I metabolism.

References

- Albersheim, P., Neukom, H., and Stutz, E. (1958) *Adv. Enzymol. Relat. Areas Mol. Biol.* **20**, 341-382
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403-410
- Barras, F., Van Gijsegem, F., and Chatterjee, A. K. (1994) *Annu. Rev. Phytopathol.* **32**, 201-234
- Bauer, D. W., Bogdanove, A. J., Beer, S. V., and Collmer, A. (1994) *Mol. Plant Microbe Interact.* **7**, 573-581
- Boulnois, G. J. (1992) *J. Gen. Microbiol.* **138**, 259-259
- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905-921
- Coutinho, P. M., and Henrissat, B. (1999) In *Recent Advances in Carbohydrate Bioengineering*, Gilbert HJ, Davies G, Henrissat B, Svensson, B (eds) pp 3-12, The Royal Society of Chemistry, Cambridge
- De Vries, R. P., Parenicova, L., Hinz, S. W., Kester, H. C., Beldman, G., Benen, J. A., and Visser, J. (2002) *Eur. J. Biochem.* **269**, 4985-4993
- Hashimoto, W., Kobayashi, E., Nankai, H., Sato, N., Miya, T., Kawai, S., and Murata, K. (1999) *Arch. Biochem. Biophys.* **368**, 367-374
- Hugouvieux-Cotte-Pattat, N., Condemine, G., Nasser, W., and Reverchon, S. (1996) *Annu. Rev. Microbiol.* **50**, 213-257
- Hugouvieux-Cotte-Pattat, N. (2004) *Mol. Microbiol.* **51**, 1361-1374
- Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24**, 946-950.
- Linker, A., Meyer, K., and Weissmann, B. (1955) *J. Biol. Chem.* **213**, 237-248
- MacGee, J., and Doudoroff, M. (1954) *J. Biol. Chem.* **210**, 617-626
- McKie, V. A., Vincken J. P., Voragen, A. G., Van Den Broek, L. E., Stimson, E., and Gilbert, H. J. (2001) *Biochem. J.* **355**, 167-177
- McNeil, M., Darvill, A. G., Fry, S. C., and Albersheim, P. (1984) *Annu. Rev. Biochem.* **53**, 625-663

- Merrit, E. A., and Murphy, M. E. P. (1994) *Acta Crystallogr. D Biol. Crystallogr.* **50**, 869-873
- Molgaard, A., Kauppinen, S., and Larsen, S. (2000) *Structure Fold. Des.* **15**, 373-383
- O'Neill, M. A., Warrenfeltz, D., Kates, K., Pellerin, P., Doco, T., Darvill, A. G., and Albersheim, P. (1996) *J. Biol. Chem.* **271**, 22923-22930
- Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276**, 307-326
- Pages, S., Valette, O., Abdou, L., Belaich, A., and Belaich, J. P. (2003) *J. Bacteriol.* **185**, 4727-4733
- Sakamoto, T., Yamada, M., Kawasaki, H., and Sakai, T. (1997) *Eur. J. Biochem.* **245**, 708-714
- Salmond, G P. C. (1994) *Annu. Rev. Phytopathol.* **32**, 181-200
- Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) In *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Sawitzky, D. (1996) *Med. Microbiol. Immunol. (Berl.)* **184**, 155-161
- Schols, H. A., Vierhuis, E., Bakx, E. J., and Voragen, A. G (1995) *Carbohydr. Res.* **275**, 343-360
- Sheldrick, G M., and Schneider, T. R. (1997) *Methods Enzymol.* **277**, 319-343
- Shevchik, V. E., Condemine, G., Robert-Baudouy, J., and Hugouvieux-Cotte-Pattat, N. (1999) *J. Bacteriol.* **181**, 3912-3919
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673-4680
- Weissbach, A., and Hurwitz, J. (1959) *J. Biol. Chem.* **234**, 705-709
- Zhang, R., Minh, T., Lezondra, L., Korolev, S., Moy, S. F., Collart, F., and Joachimiak, A. (2005) *Proteins* **60**, 561-565

CHAPTER III

A NOVEL ENZYME REACTION CATALYZED BY UNSATURATED GLYCOSIDE HYDROLASES

Section 1

Structure of Unsaturated Glucuronyl Hydrolase Complexed with Substrate

UGL is peculiar in that the enzyme acts on unsaturated saccharides but not on usual saccharides, i.e., saturated saccharides. This suggests that the enzyme catalyzes a novel reaction distinct from general glycosidases.

In Chapter I, the crystal structure of UGL (PDB, Accession No.: 1VD5) was determined and showed that it includes a α_6/α_6 -barrel topology, which is found in the α/α -toroidal fold in the six-hairpin glycosidase superfamily. In glycosidases, two carboxyl and carboxylate groups are reported to generally function as catalytic residues, and based on structural and mutagenesis studies, two candidate residues (Asp-88 and Asp-149) were proposed to be located in the pocket. The catalytic reaction mechanism of the enzyme remains to be elucidated, however. In this section, the catalytic reaction mechanism of UGL was elucidated through the determination of the crystal structure of a mutant UGL and its complex with a substrate (Δ GlcA-GalNAc).

Materials and Methods

Crystallization and X-ray diffraction

Wild-type and a UGL mutant having Asp-88 substituted with Asn (D88N) were overexpressed in *Escherichia coli* and purified. The purified enzymes were concentrated by vivaspin 0.5 ml concentrators with 10,000 molecular weight cut off (MWCO) membranes (Vivascience, Hannover, Germany) to a final concentration of 10 mg/ml. The concentrated enzymes,

in 20 mM potassium phosphate (pH 7.0),

were then used in the crystallization step.

Crystals were obtained by using hanging

drop vapor diffusion and microseeding.

Hanging drops were prepared by mixing

3 μ l of the enzyme solution with an

equal volume of reservoir solution

containing 40% (w/v) polyethylene

glycol 10000 and 0.15 M Tris-HCl (pH

8.6), and equilibrated with 0.5 ml of the

reservoir solution at 20°C. After two

weeks, minute crystals formed in drops.

These initial crystals were crushed by a

stainless steel needle and transferred to

hanging drops equilibrated under

conditions similar to those used for initial crystals, except for 25% (w/v) polyethylene glycol 10000.

Rod-shaped crystals grew to a maximum $0.1 \times 0.1 \times 0.5$ mm during the three months after seeding.

One D88N crystal was soaked in the substrate solution containing 400 mM Δ GlcA-GalNAc

(chondroitin disaccharide di-OS sodium salt; Sigma, St. Louis, MO, USA), 25% (w/v) polyethylene

glycol 10000, and 0.15 M Tris-HCl (pH 7.5) for 10 min at 20°C. The pH optimum of the enzyme is

about pH 6. pK_{a2} is about pH 7.6. Crystals used here were therefore moved into the solution

containing Tris pH 7.5. Crystals were placed in a cold nitrogen gas stream at -173°C . X-ray

diffraction images of crystals were collected using a JUPITER 210 CCD detector (Rigaku, Tokyo,

Japan) with synchrotron radiation at a wavelength of 0.80 (D88N) or 1.00 Å

(D88N/ Δ GlcA-GalNAc complex and wild-type) at the BL-38B1 station of SPring-8 (Hyogo, Japan).

Table 1. Data collection and refinement statistics

	Wild-type	D88N	D88N/ Δ GlcA-GalNAc
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell parameters (Å)	a = 88.0, b = 95.3, c = 95.3	a = 87.9, b = 95.4, c = 95.3	a = 87.9, b = 95.0, c = 95.4
Data collection			
Resolution limit (last shell) ^a (Å)	50.0–1.60 (1.66–1.60)	50.0–1.52 (1.57–1.52)	50.0–1.90 (1.97–1.90)
Measured reflections	403641 (33327)	631135 (41083)	310585 (9352)
Unique reflections	93414 (10099)	119866 (11412)	58182 (4251)
Redundancy	4.3 (3.3)	5.3 (3.6)	5.3 (2.2)
Completeness ($ I > \sigma(I)$) (%)	87.7 (96.1)	97.2 (93.6)	91.3 (67.7)
$I/\sigma(I)$	16.4 (2.3)	16.5 (1.2)	7.9 (1.6)
R_{merge} (%) ^b	7.9 (27.6)	6.5 (39.2)	8.8 (24.9)
Refinement			
Final model	754 (377 \times 2) residues, 835 water molecules	754 (377 \times 2) residues, 866 water molecules	754 (377 \times 2) residues, 766 water molecules, 2 Δ GlcA-GalNAc
Resolution limit (Å)	50.0–1.60 (1.67–1.60)	50.0–1.52 (1.57–1.52)	50.0–1.90 (1.97–1.90)
Used reflections	93134 (10078)	119790 (11142)	58150 (4260)
Completeness ($ I > \sigma(I)$) (%)	87.7 (96.2)	97.0 (91.4)	91.4 (67.9)
Average B-factor (\AA^2)			
Protein (A-molecule, B-molecule)	13.0, 16.8	13.9, 17.1	16.5, 20.4
Δ GlcA-GalNAc (A) (Δ GlcA, GalNAc)			20.5, 37.6
Δ GlcA-GalNAc (B) (Δ GlcA, GalNAc)			37.5, 57.3
Water	25.7	26.3	27.8
R-factor (%) ^c	18.8 (25.5)	18.7 (30.0)	18.5 (29.0)
R_{free} (%) ^d	22.1 (28.1)	21.2 (30.5)	23.5 (32.2)
Root-mean-square deviations			
Bond (Å)	0.005	0.005	0.005
Angle (deg)	1.21	1.20	1.20
Ramachandran plot (%)			
Most-favored regions	88.7	89.6	88.7
Additionally-allowed regions	11.0	10.3	11.0
Generously-allowed regions	0.3	0.1	0.3

^a Data in highest resolution shells is given in parentheses.

^b $R_{\text{merge}} = \sum |I_i - \langle I \rangle| / \sum I_i \times 100$, where I_i is the intensity of individual reflection and $\langle I \rangle$ is the mean intensity of all reflections.

^c R-factor = $\sum |F_o - F_c| / \sum |F_o| \times 100$, where F_o is the observed structure factor and F_c is the calculated structure factor.

^d R_{free} was calculated from 10% of reflections randomly selected, as defined by the CNS.

Images were processed using DENZO and SCALEPACK software (Otwinowski & Minor, 1997) to a resolution of 1.60, 1.52, or 1.90 Å (Table 1).

Structure determination and refinement

Structures were determined using molecular replacement and refined using CNS program ver. 1.1 (Brünger *et al.*, 1998) with the UGL structure (1VD5) as the reference model. Several rounds of positional and *B*-factor refinement were conducted, followed by manual model building using the TURBO-FRODO program (AFMB-CNRS, Marseille, France) to improve the model by increasing data to a resolution of 1.60, 1.52, or 1.90 Å for wild-type, D88N, and D88N/ΔGlcA-GalNAc with the CNS program (Table 1). The stereoquality of the model was assessed using the PROCHECK (Laskowski *et al.*, 1993) and WHAT-CHECK (Hooft *et al.*, 1996) programs. Structural alignment was conducted by superimposition using a fitting program in TURBO-FRODO. Ribbon plots were prepared using the MOLSCRIPT (Kraulis, 1991), Raster3D (Merrit & Murphy, 1994), and GRASP (Nicholls *et al.*, 1991) programs. Carbohydrate conformations (puckering parameters) as defined by Cremer and Pople (1975) were analyzed using the PLATON program (Speck, 2001).

Site-directed mutagenesis

Arg-221, His-193, or Gln-211 in UGL was replaced with an alanine residue by using a QuickChange site-directed mutagenesis kit (Stratagene Co., CA, USA); mutations were confirmed by DNA sequencing. The plasmid pET3a-UGL, which is the expression vector for wild-type UGL (Hashimoto *et al.*, 1999), was used as a template. The following primers were used: R221A (Arg-221 → Ala-221), 5'-AGCACGTGGACGGCCGGCCAGGCTTGG-3' and 5'-CCAAGCCTGGCCGGCCGTCCACGTGCT-3'; H193A (His-193 → Ala-193), 5'-CGACGATTCGAGCTATGCGACGTTCTACTTCGACC-3' and 5'-GGTCGAAGTAGAACGTTCGCATAGCTCGAATCGTTCG-3'; and Q211A (Gln-211 → Ala-211), 5'-GGCGGCACGCACGCGGGCAACACCGAC-3' and 5'-

GTCGGTGTGCCCC**G**CGTGCGTGCCGCC-3' (mutations are indicated by bold letters). Cells of *E. coli* BL21(DE3) were transformed with mutant plasmids, i.e., pET3a-UGL(R221A), pET3a-UGL(H193A), and pET3a-UGL(Q211A). Mutant enzymes were expressed and purified using a procedure similar to that used for wild-type UGL. Enzyme purity was assessed by SDS-PAGE followed by Coomassie Brilliant Blue R250 staining.

Enzyme assay

Enzyme-substrate reactions for the wild-type enzyme and mutants were conducted at 30° C as follows: the reaction mixture consisted of 50 mM sodium phosphate (pH 6.5), the substrate, and enzyme in a 500 µl reaction volume. Since a large amount of glycosaminoglycan disaccharides was unavailable, gellan tetrasaccharide (Δ GlcA-Glc-Rha-Glc) produced from gellan by gellan lyase was used as the UGL substrate. Enzyme activity was measured by monitoring the decrease in absorbance at 235 nm, corresponding to the loss of the C=C double bond of the substrate (Fig. 1d, Chapter I). k_{cat} and K_m parameters were determined by nonlinear fitting to the Michaelis–Menten equation. Enzyme concentration was determined by UV spectrophotometry by using theoretical molar extinction coefficient $\epsilon_{280} = 99,570 \text{ (M}^{-1}\text{cm}^{-1}\text{)}$. The dependence of k_{cat}/K_m on pL (pH or pD) was measured in solutions consisting of 50 mM acetate pL 3.5 – 5.5 or sodium phosphate pL 5.2 – 8.2. pD values were estimated by adding 0.41 to the pH meter reading (Convington *et al.*, 1968). The dependence of k_{cat}/K_m on pL was fitted to the following equation: $(k_{cat}/K_m)^{obs} = (k_{cat}/K_m)^0 / (1 + 10^{pK_{a1}-pL} + 10^{pL-pK_{a2}})$. The solvent kinetic isotope effect (SKIE) on k_{cat} and K_m was measured near the optimum pL (pH 6.0 or pD 6.6).

Gas chromatography/mass spectrometry (GC/MS) analysis

Enzyme reaction mixtures [2 µmol sodium phosphate (pH 6.5), 2.5 mg Δ GlcA-GalNAc, and 30 µg UGL in a 100 µl reaction volume with 88, 50, 0% ^{18}O water (18O, 95%, Cambridge Isotope Laboratories, MA, USA)] were incubated at 30°C overnight. Enzyme molecules in reaction mixtures were removed by vivaspin 0.5 ml concentrators with 10,000 MWCO membranes.

Reduction, methylation, and GC/MS analysis were conducted based on the method of Chai *et al.* (1995) as follows: After filtrated products (10 μ l) were lyophilized and incubated with 0.5 ml of 1% NaBH₄ at room temperature overnight, excess NaBH₄ was decomposed by 50% acetic acid. Reduced products were purified five times with 100% methanol. Methylation was conducted by the NaOH-CH₃I method with reduced products (Ciucanu & Kerek, 1984). Methylated products were extracted with chloroform, dried with Na₂SO₄, and redissolved in chloroform for GC/MS analysis. GC/MS analysis was conducted on a Hewlett Packard HP5890A gas chromatograph (Agilent Technologies, CA, USA) with a JMS-DX303 mass spectrometer (Jeol, Tokyo, Japan) and a Supelco SPB-5 capillary column (31 m x 0.25 mm i.d. x 0.25 μ m film thickness; Sigma-Aldrich Corporate, MO, USA) with He as the carrier gas. The initial column temperature of 60°C was held for 1 min and increased to 280°C at 8° C/min. Mass spectra (m/z 35-500) were acquired at 70 eV, 300 μ A, and a source temperature of 250°C.

Results and Discussion

Structure determination

UGL of *Bacillus* sp. GL1 is a monomeric enzyme with a molecular mass of approximately 43 kDa (377 amino acid residues) (Hashimoto *et al.*, 1999). The hexagonal crystal structure of wild-type UGL has been described in Chapter I. This structure contains one glycine and two DTT molecules derived from the crystallization solution in the deep pocket (1VD5), making it difficult to determine the structure of the enzyme-substrate complex. Asp-88 is located in the deep pocket, and is completely conserved in UGL and its homologues (Fig. 2, Introduction). The molecular activity ($k_{cat} = 0.00057 \text{ s}^{-1}$) of mutant D88N is about 13000-fold lower than that ($k_{cat} = 7.3 \text{ s}^{-1}$) of the wild-type enzyme, although the Michaelis constants (K_m) of both are comparable. Due to the lack of enzyme activity, a mutant having an affinity for the substrate has a greater advantage in forming the enzyme-substrate complex than the wild-type enzyme. Orthorhombic crystals (wild-type and

D88N) were obtained by hanging drop vapor diffusion and microseeding by using polyethylene glycol 10000 as a precipitant.

Δ GlcA-GalNAc produced from chondroitin by chondroitin lyase is a disaccharide and a good substrate for UGL (Fig. 1a, Introduction). The structure of the D88N/substrate complex was first determined using Δ GlcA-GalNAc (D88N/ Δ GlcA-GalNAc). The results of data collection for structures using synchrotron radiation are summarized in Table 1.

Refined models consist of 754 (377×2 monomers) amino acid residues and 2 Δ GlcA-GalNAc molecules (Fig. 1a). Two identical monomers are in an asymmetric unit, designated A- and B-molecules for convenience. The two Δ GlcA-GalNAc models are well fitted, but the average *B*-factor of the B-molecule is higher than that of the A-molecule (Table 1). The Δ GlcA-GalNAc (A) is better characterized than the B-molecule. Δ GlcA (A) is particularly well settled (Fig. 2b).

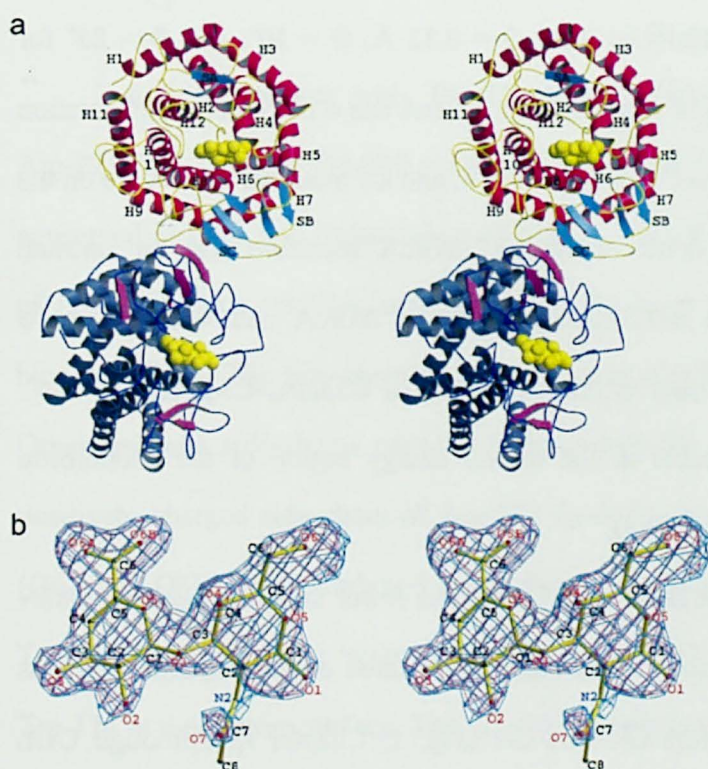


Fig. 1. Crystal structure of UGL complexed with substrate. (a) Overall structure of UGL bound with substrates. Two α/α -barrels are shown in red (A-molecule) and gray (B-molecule). Two yellow CPK models indicate Δ GlcA-GalNAc. Other colors denote secondary structure elements (cyan and magenta, β -strands; yellow and blue, loops). (b) Electron density of the Δ GlcA-GalNAc substrate in the omit ($F_o - F_c$) map calculated without the substrate and contoured at the 2.5 σ (magenta) and 3.0 σ (cyan) levels. Characters indicate atom names of the substrate.

Structures of orthorhombic crystals (wild-type, D88N, and D88N/ Δ GlcA-GalNAc)

coincide, except for an active site, described later. The structures are very similar to the hexagonal crystal structure of the wild-type enzyme, showing no significant difference from the overall structure. The barrel structure forms a deep pocket consisting of long loops and β -sheets, where the substrate is bound in D88N/ Δ GlcA-GalNAc. The deep pocket, approximately ~ 20 Å in diameter at the lip and ~ 15 Å in depth, is strongly suggested to be an active site.

Δ GlcA-GalNAc bound at the active site

The structure of the enzyme-substrate complex contains one substrate, the Δ GlcA-GalNAc molecule, per monomer. Two planes of sugar rings are almost parallel with the axis of symmetry of the α_6/α_6 -barrel structure (Figs. 1a and 2). The substrate is bound at the bottom of the deep pocket, indicating that it occupies two subsites, -1 and +1, at the active site. The nomenclature of subsites complies with that described by Davies *et al.* (1997). Δ GlcA-GalNAc is bound at the surface surrounded by aromatic amino acid residues (Phe, Trp, and Tyr) (Fig. 2). The puckering parameters (Cremer & Pople, 1975) of bound Δ GlcA-GalNAc were $Q = 0.52$ Å, $\Theta = 50^\circ$, and $\Phi = 85^\circ$ for Δ GlcA, and $Q = 0.59$ Å, $\Theta = 4^\circ$, and $\Phi = 354^\circ$ for GalNAc. Δ GlcA has a half-chair configuration (2H_1) and GlcNAc a stable chair configuration (1C_4). The C3, C4, and C5 atoms of Δ GlcA are on the same plane due to the C4=C5 double bond restriction. Δ GlcA therefore has an unusual configuration. The dihedral angles of the α -1,3-linkage between Δ GlcA and GalNAc are φ ($O5_{\Delta\text{GlcA}}-C1_{\Delta\text{GlcA}}-O3_{\text{GalNAc}}-C3_{\text{GalNAc}}$) = -74° and ψ ($C1_{\Delta\text{GlcA}}-O3_{\text{GalNAc}}-C3_{\text{GalNAc}}-C3_{\text{GalNAc}}$) = -146° . These torsion angles are preferable and stable in the lowest energy region of the chondroitin disaccharide isoenergy map (Zsiska & Meyer, 1993).

The number of interactions between D88N and GalNAc is fewer than that between D88N and Δ GlcA (Fig. 2b and Table 2). UGL also acts on other unsaturated oligosaccharides such as Δ GlcA-GlcNAc, Δ GlcA-Man-Glc, and Δ GlcA-Glc-Rha-Glc (Fig. 1, Chapter I). Although UGL activity toward these substrates differs to some extent, all of these substrates contain Δ GlcA at the nonreducing terminus. It appears that subsite +1 can accommodate various saccharides. In other

words, UGL strictly recognizes Δ GlcA at subsite -1 rather than the saccharide at subsite +1. This is supported by the fact that the average *B*-factor of Δ GlcA is significantly lower than that of GalNAc (Table 1).

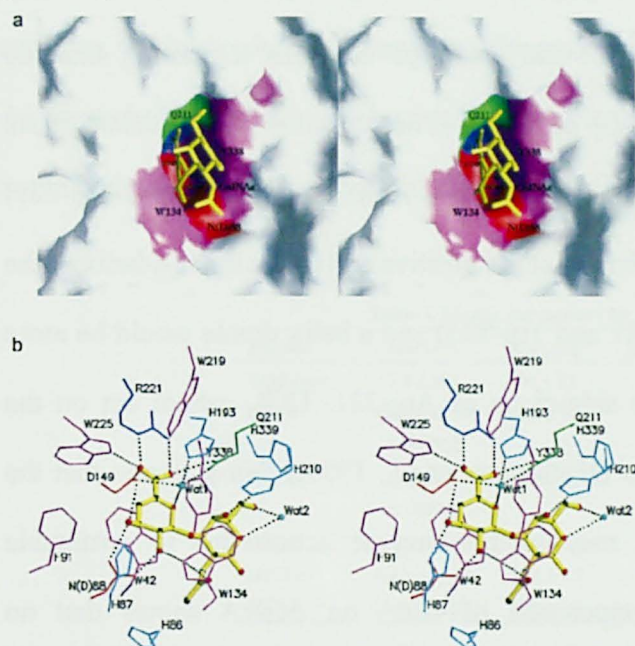


Fig. 2. The structure of UGL bound with the substrate at the active site. (a) The structure is represented as a white molecular surface model (magenta, Trp-42, Trp-134, and Tyr-338; red, Asn-88 and Asp-149; blue, Arg-221; green, Gln-211). The bound substrate (Δ GlcA-GalNAc) is represented by yellow bond models. (b) Residues interacting with the substrate are represented by bond models. Side-chains are indicated in red (Asp and Asn), blue (Arg), purple (Trp, Tyr, Phe), cyan (His), and green (Gln). The substrate is denoted by a yellow bond model (oxygen atom, red; carbon atom, black; nitrogen atom, blue). Two water molecules (Wat-1 and Wat-2) are represented as light blue balls. Hydrogen bonds are shown as dashed lines.

Δ GlcA interacts with Trp-42, Asn-88, Phe-91, Trp-134, Asp-149, Gln-211, Trp-219, Arg-221, Trp-225, and Tyr-338 residues (Fig. 2b and Table 2). All of these residues are completely conserved in UGL and its homologues (Fig. 2, Introduction). In particular, the aromatic side-chain of Trp-42 shows a stacking interaction with the sugar ring of Δ GlcA and plays an essential role in binding Δ GlcA. The side-chain of Trp-225 forms a hydrogen bond to the carboxyl group of Δ GlcA. Other aromatic side-chains partially interact with the Δ GlcA pyranose ring hydrophobically. The positively charged side-chain of Arg-221 faces the carboxyl groups of Δ GlcA. The two O atoms (O6A and O6B) of the carboxyl group form hydrogen bonds directly with Arg-221, Gln-211, and Trp-225. Through direct hydrogen bonds, O2 and O3 atoms associate with Asn-88, Asp-149, and Trp-134 at the bottom surface. The O5 atom interacts with Asp-149, His-193, and Gln-211 through a water molecule (Wat-1). For some polysaccharide lyases such as chondroitin lyase (Lunin *et al*, 2004), hyaluronate lyase (Jedrzejewski *et al.*, 2002), and xanthan lyase (Maruyama *et al.*, 2005), the

asparagine residue is close to the carboxyl group and is important for catalysis because it acts to neutralize the negative charge of the carboxyl group. Thus, to determine the effect of Arg-221 on the substrate, the mutant enzyme, R221A, was prepared and assayed (Table 3). The kinetic parameters of R221A are comparable to those of the wild-type enzyme, suggesting that the interaction between Arg-221 and the carboxyl group of Δ GlcA is not essential to bind the substrate or to participate in catalysis. Other than Arg-221, there are two residues (Gln-211 and Trp-225) forming hydrogen bonds to this group. Δ GlcA is bound at the positive end of the inner α -helix of the α_6/α_6 -barrel, suggesting that two residues (Gln-211 and Trp-225) and a helix dipole would be more effective for neutralizing the substrate than the side-chain of Arg-221. UGL cannot act on the substrate with GlcA at the nonreducing terminus (Hashimoto *et al.*, 1999); this suggests that the presence of the C4 hydroxyl group of GlcA may inhibit enzyme action due to unsuitable accommodation at subsite -1, but the superimposition of GlcA on Δ GlcA shows that no three-dimensional obstacle particularly surrounds the C4 hydroxyl group. The possible reason for the inability of UGL to act on usual glucuronyl saccharides is discussed later.

Table 2. D88N/ Δ GlcA-GalNAc and UGL interaction at the active site

Hydrogen bonds (<3.3 Å)				van der Waals contact (C-C distance <4.5 Å)			
Sugar	Atom	Protein atom	Distance (Å)	Sugar	Atom	Protein atom	
Δ GlcA	O2	Asn-88	O61	2.5	Δ GlcA	C1	Trp-42 C61, C62, C62, Cy, C62
		Trp-134	N61	2.9			Trp-134 C62
	O3	Asn-88	N62	3.0		C2	Trp-134 C62, C62
		Asp-149	O52	3.0			Asn-88 Cy
	O5	Wat-1		3.2		C3	Trp-42 C62, C62, C63, C62, C62, C63
	O6A	Arg-221	N6	2.7			Asn-88 Cy
		Trp-225	N61	3.3			Phe-91 C61
	O6B	Gln-211	N62	3.2			Asp-149 C61, Cy
		Arg-221	N62	2.9		C4	Trp-42 C62, C62, C63, C62, C62, C63
							Asp-149 C61, Cy
GalNAc	O4	His-210	N62	3.0		C5	Trp-42 C62, C62, C62, C62, C63
	O5	Wat-2		3.3			Trp-42 C62, C62, C62
	O6	Wat-2		2.8			Trp-219 C62, C62
	O7	Trp-134	N61	3.3			Arg-221 C61
							Tyr-338 C61

GalNAc interacts with Trp-42, His-86, Trp-134, His-210, Tyr-338, and His-339 residues (Fig. 2b and Table 2). The *N*-acetyl group of GalNAc (N2, C7, C8, and O7 atoms) has few interactions with the enzyme and is thus disordered (Fig. 2b and Table 2). Another substrate for

UGL has Man or Glc residue at subsite +1, which contains no such group (Fig. 1, Introduction). Two residues (His-210 and His-339) in UGL are replaced with arginine and serine in other homologues (Fig. 2, Introduction). Based on the broad specificity of subsite +1 and conserved residues, the hydrophobic moiety, e.g., the pyranose ring, is important for accommodation at subsite +1 through hydrophobic interactions with aromatic side-chains of Tyr-338 and Trp-134, although some hydrogen bonds are present in binding GalNAc.

Table 3. Kinetic parameters for wild-type and mutant UGLs

Enzyme	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($\mu M^{-1}ms^{-1}$)	Relative k_{cat}/K_m (%)
Wild-type	7.3 ± 0.3	90 ± 11	81 ± 27	100
D88N	0.00057 ± 0.000052	200 ± 38	0.0029 ± 0.014	0.0036
D149N	0.0059 ± 0.00018	60 ± 5.8	0.098 ± 0.031	0.12
R221A	4.2 ± 0.3	69 ± 12	61 ± 24	75
H193A	0.42 ± 0.012	980 ± 47	0.43 ± 0.25	0.53
Q211A	0.20 ± 0.015	170 ± 26	1.1 ± 0.57	1.4

Comparison to the substrate-free structure

The overall structures of the wild-type enzyme (hexagonal crystal form and orthorhombic one), D88N, and D88N/ Δ GlcA-GalNAc are essentially identical as stated above. In the deep pocket, they share structures almost similar except for the side-chains of Asp-149 and Trp-134, which move based on ligand binding (Fig. 3). Although the side-chain of Asp-149 was modeled with two alternate conformations in the wild-type/apo form (Fig. 3a, wild-type model), it is fixed with the ligand in the holo form by rotating about 85° (Fig. 3, D88N/ Δ GlcA-GalNAc and Fig. 3b, hexagonal wild-type models). The side-chain of Trp-134 is also fixed based on ligand binding by rotating about 110° . Although apo forms (orthorhombic wild-type and D88N) have more water molecules at the ligand than the complex form (D88N/ Δ GlcA-GalNAc and hexagonal wild-type), the Wat-1 molecule is located at the same position in all structures (Figs. 2b and 3). The hexagonal wild-type structure has one additional glycine residue, two DTT molecules derived from the crystallization

solution, and two water molecules at the active site (Fig. 3b). The location of the carboxyl group of the glycine molecule in the wild-type enzyme corresponds to that of Δ GlcA in substrate-bound D88N, and both carboxyl groups interact with the enzyme identically. Two hydroxyl groups (O2 and O3) of Δ GlcA that form hydrogen bonds with Asn-88 are replaced with two additional water molecules bound to Asp-88 through hydrogen bonds. One DTT molecule is located at subsite +1 corresponding to the position of GalNAc, and has hydrophobic interactions with Tyr-338. The second DTT molecule interacts mainly with the aromatic side-chain of Trp-134.

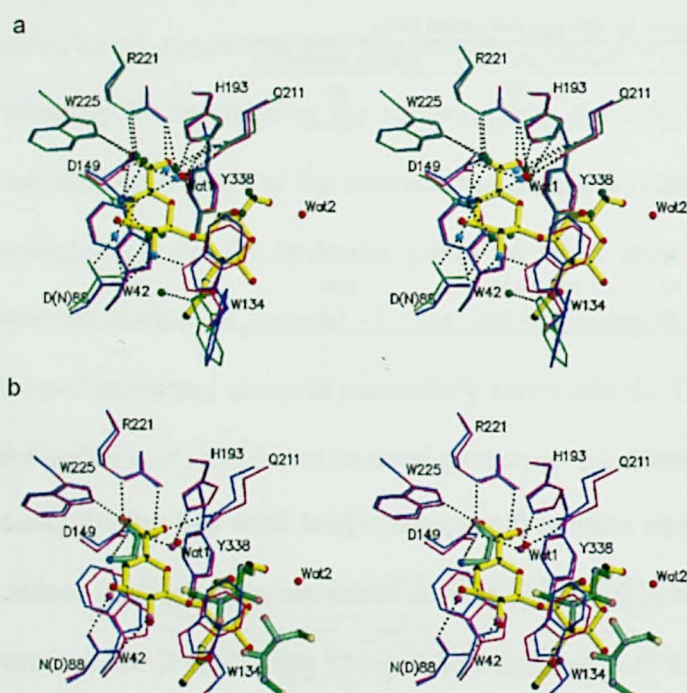


Fig. 3. Active site structure. (a) Structural comparison of wild-type, D88N, and D88N/ Δ GlcA-GalNAc complex structures at the active site. Amino acid residues of wild-type are shown in blue, D88N in green, and D88N/ Δ GlcA-GalNAc in magenta. The substrate (Δ GlcA-GalNAc) structure is denoted by a yellow bond model (oxygen atom, red; carbon atom, black; nitrogen atom, blue), similar to Fig. 2. Water molecules of wild-type (cyan), D88N (deep green), and D88N/ Δ GlcA-GalNAc (red) are represented as ball models. Hydrogen bonds in wild-type and D88N are shown as dashed lines. (b) Structural comparison of the D88N/ Δ GlcA-GalNAc complex and wild-type/glycin-DTT complex structures at the active site. The wild-type complex structure is indicated in blue. Glycine and DTT molecules of the wild-type complex structure are denoted by green bond models (oxygen atom, red; carbon atom, black; nitrogen atom, blue; sulfur atom, yellow). Water molecules of the wild-type complex are shown as purple ball models. Hydrogen bonds in the wild-type complex structure are shown as dashed lines.

Catalytic mechanism

UGL acts strictly on unsaturated oligosaccharides with Δ GlcA at the nonreducing terminus, and is inert on saccharides containing GlcA. Products obtained through the enzyme reaction are 4-deoxy-L-threo-5-hexosulose-uronate and the leaving saccharide with a new nonreducing terminus (Fig. 1d, Introduction). General β -glycosidase cannot act on unsaturated saccharides with Δ GlcA at the nonreducing terminus. The reaction mechanism of the hydrolysis of these unsaturated saccharides thus remains unclear.

The structure of UGL in a complex with the substrate (D88N/ Δ GlcA-GalNAc) determined

in the present study may provide valuable clues to the nature of the catalytic reaction mechanism. Regarding the binding of Δ GlcA to the active site, no atom is located close to glycosidic oxygen (O3 atom of GalNAc) within the hydrogen bond distance (<3.5 Å). The characteristic environment of the active site is as follows (Fig. 2b): a water molecule (Wat-1) is present at the “ α -face position” of the Δ GlcA pyranose ring and forms four hydrogen bonds with Asp-149 (3.1 Å), His-193 (2.7 Å), Gln-211 (2.6 Å), and O5 atom (3.2 Å) of Δ GlcA, but not with glycosidic oxygen (4.2 Å); Wat-1 is also located close to the C5 carbon atom (3.4 Å) but not to the C1 carbon atom (3.9 Å); Asp-149 is close to the C4 atom (2.9 Å) but not to glycosidic oxygen (5.8 Å) and C1 carbon (4.8 Å) atoms; Asn-88 located at the mutation site (wild-type, Asp-88) is also far from glycosidic oxygen (4.9 Å) and C1 carbon (4.3 Å) atoms; no water molecule exists at the “ β -face position” of the pyranose ring; Δ GlcA has an alkene C4=C5 double bond, enriched in electrons and termed an “electron sink.” These characteristics are unusual in glycoside hydrolases (Davies & Henrissat, 1995; Rye & Withers, 2000). In particular, no hydrogen bonds are formed with glycosidic oxygen. A catalytic reaction mechanism of UGL is therefore postulated as follows; the enzyme catalyzes a water addition reaction (hydration) of the vinyl ether group (C4=C5–O5) and subsequent hydrolysis of the glycosidic bond in Δ GlcA. This postulation is based on the structure of the enzyme-substrate complex, the characteristic of the substrate [i.e., susceptibility of the vinyl ether group with an “electron sink” in Δ GlcA to the water addition reaction (Aoyama *et al.*, 2004; Kresge, 1987)], enzyme properties determined through kinetics on site-directed mutants as described later. A possible catalytic reaction mechanism of UGL is shown in Fig. 4. Asp-149 acts as a general acid catalyst and donates a proton to the double bond (C4 atom). Asp-149 acts as a general base catalyst and deprotonates the water molecule (Wat-1); the deprotonated water molecule then attacks the C5 atom of Δ GlcA. The product hemiketal is unstable and readily converts to α -keto acid (hemiacetal). Due to aldehyde-hemiacetal equilibrium, the resultant hemiacetal finally converts to the aldehyde (4-deoxy-L-threo-5-hexosulose-uronate) and the leaving saccharide through cleavage of the

glycosidic bond. These consecutive reactions appear to be similar to the oxymercuration of unsaturated hyaluronate disaccharide Δ GlcA-GlcNAc (Ludwigs *et al.*, 1987). In other words, the water addition reaction (hydration) of the vinyl ether group in Δ GlcA and spontaneous conversion of hemiketal to aldehyde and the leaving saccharide via α -keto acid (hemiacetal) are also demonstrated in the reaction mixture of Δ GlcA-GlcNAc and mercuric salts.

Many glycosidases hydrate C1 anomeric configuration-free unsaturated substrates (D-glycals, C2=C1-O5 or D-glyco-enitols, O5-C1=C). Their catalytic mechanism and anomeric configuration of products are regulated by their own specificities in hydrolysis, such as “retaining” or “inverting mechanisms” (Lehmann & Schröter, 1972; Schlesselmann *et al.*, 1982; Weiser *et al.*, 1988). UDP-GlcNAc 2-epimerase catalyzes the epimerization of UDP-GlcNAc and UDP-*N*-acetylmannosamine (UDP-ManNAc) in bacteria, or the conversion of ManNAc and UDP in mammals. Their reactions involve elimination to generate a vinyl ether group (2-acetamidoglucal: C2=C1-O5) and the readdition of UDP in bacteria or a water molecule in mammals (Morgan *et al.*, 1997; Chou *et al.*, 2003). The proposed mechanism of UGL is similar. In particular, the carboxylate group (Asp-149: UGL) acting as the general acid/base catalyst to protonate the vinyl group and deprotonate the catalytic water molecule is equivalent to those of retaining glycosidases. The reaction mechanism consisting of the hydration of the vinyl ether group (C4=C5-O5) triggering hydrolysis of the glycosidic bond and the arrangement of the active site residue, however, are novel and intriguing (Figs. 2b and 4).

The location of Asp-149 is suitable for the protonation of the C4 atom of Δ GlcA (Fig. 2b), and Asp-149 is completely conserved in UGL and its homologues. A mutant enzyme D149N shows significantly reduced molecular activity (0.12%) when compared to the wild-type enzyme, as described previously (Table 3). To appropriately function as a general acid catalyst and donate a proton to the C4 atom of Δ GlcA, Asp-149 must be protonated (high pKa). The environment of Asp-149 is polar and there are few hydrogen bonds with Asp-149 in the complex structure (Fig. 2b).

They would serve to elevate its pKa. This side-chain accepts its own NH backbone group, being essential for properly orienting this side-chain for catalysis.

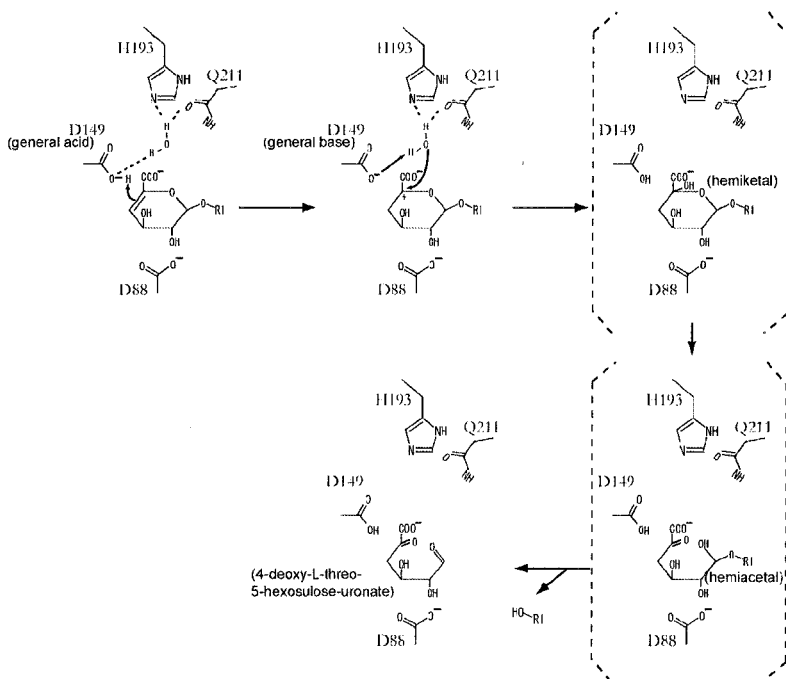


Fig. 4. Proposed catalytic reaction mechanism of UGL. The catalytic reaction proceeds through the water addition reaction of the vinyl ether group, as described in the text. Important residues surrounding the substrate are shown.

His-193, which is completely conserved in UGL and its homologues, is also located close to Δ GlcA and has an ionizable side-chain (Fig. 2b). To determine the possibility of His-193 as a catalytic residue, a mutant enzyme (H193A) was prepared and assayed (Table 3). k_{cat} of H193A (0.42 s^{-1}) is lower than that of the wild-type enzyme (7.3 s^{-1}); but significantly higher than that of D149N (0.0059 s^{-1}). K_m of H193A ($980 \text{ }\mu\text{M}$) is considerably higher than that of the wild-type enzyme ($90 \text{ }\mu\text{M}$) and D149N ($60 \text{ }\mu\text{M}$). The difference in K_m between H193A and the wild-type enzyme is thought to be due to the conformational change of the active site. Kinetic parameters for D149N and H193A suggest that Asp-149 is the most conceivable general acid/base catalyst among residues located nearest to Δ GlcA. Since k_{cat} of mutants H193A and Q211A is lower than that of the wild-type enzyme, His-193 and Gln-211 probably play a crucial role in fixing Wat-1 at a suitable position for catalysis (Table 3).

Asp-88 is also completely conserved in UGL and its homologues, and mutant enzyme D88N is almost inactive (0.0036% activity of the wild-type enzyme) and less active than D149N

(0.12%) (Table 3). Asp-88 is thought to both stabilize the carbenium ion at C5 of Δ GlcA during catalysis and mainly to tie protons of hydroxyl groups of O2 and, most importantly, O3 in hydrogen bonds (Figs. 2b and 4). Otherwise, the O3 hydroxyl group would be able to form a hydrogen bond to the side-chain of Asp-149, preventing its protonation and also directing it away from C4 and toward O3 (Fig. 2b).

The SKIE was measured to confirm the participation of the general acid and water molecule in the reaction. The SKIE on the second-order rate constant (k_{cat}/K_m) of UGL depends on the pL (pH or pD). The optimal pL (pH = 6.0 or pD = 6.6) and the two pKas ($pK_{a1} = 4.4$ or 5.2, $pK_{a2} = 7.6$ or 8.2) increased in D₂O. The shift of 0.6 units in pK_{a2} , thought to be pK_a of the side-chain of Asp-149, is typically seen for acids (Laughton & Robertson, 1969). There is a significant SKIE on both k_{cat} and k_{cat}/K_m near the optimal pL. Catalysis in H₂O (pH 6.0) showed $k_{cat} = 6.8 \pm 0.7$ (s⁻¹), $K_m = 62 \pm 13$ (μ M), and $k_{cat}/K_m = 0.11$ (μ M⁻¹s⁻¹); that in D₂O (pD 6.6) gave $k_{cat} = 3.3 \pm 0.2$ (s⁻¹), $K_m = 66 \pm 11$ (μ M), and $k_{cat}/K_m = 0.05$ (μ M⁻¹s⁻¹). The SKIE is thus 2.1 for k_{cat} and 2.2 for k_{cat}/K_m . General acid catalysis is associated with the SKIE for k_{cat} of 1.5-2.5 (Fife & Brod, 1970). These observed SKIEs are typical for general acid catalysis and indicate the mechanism involving proton transfer in the rate-determination step (Oyama & Tidwell, 1976).

The structural basis for hydration of the intrinsic substrate with the vinyl ether group through enzymatic catalysis by the asparagine residue has been reported with isochorismatase (PhzD) from *Pseudomonas aeruginosa* (Parsons *et al.*, 2003). PhzD is involved in phenazine biosynthesis and catalyzes the hydrolysis of the vinyl ether group of 2-amino-2-deoxyisochorismate. Although UGL and PhzD show no similarity in amino acid sequence or overall structure, their amino acid residues interacting with the substrate are comparable as follows: the catalytically important residue (Asp-149) in UGL is equivalent to Asp-38 in PhzD, and the other important residues (Trp-42 and Tyr-338) in UGL correspond to those of Trp-94 and Tyr-151 in PhzD. Asp-38 in PhzD is considered to be a general acid catalyst and is thought to hydrate the vinyl ether group

(Parsons *et al.*, 2003). The magnitude of k_{cat} ($\sim 10\text{ s}^{-1}$) of UGL is close to that of PhzD. This evidence suggests that both UGL and PhzD have a similar enzyme reaction mechanism (hydration), and that the protonation of the vinyl ether group by aspartate residues is the rate-limiting step of the reaction.

The detailed structural analysis of the reaction product of UGL has not been clarified, although the product is confirmed to be α -keto acid derived from Δ GlcA. GC/MS analysis on the reaction product was conducted. Fig. 5 shows the electron ionization mass spectrum of reduced and methylated Δ GlcA after the UGL reaction in 0, 50, or 88% ^{18}O water. The fragmentation pattern of this product (m/z 101, 129, 133, 141, 161, 173, and 191) was equivalent to that of reduced and methylated Δ GlcA released through the reaction of unsaturated heparin disaccharide and mercuric salts (Chai *et al.*, 1995) (Fig. 5). Only three ion peaks (m/z 105, 163 and 193) were observed, however, with the ^{18}O water hydrolysis sample, and these peak heights increased with an increasing ^{18}O water ratio (Fig. 5b, c). Further fragmentation ion peaks (m/z 73 and 131) produced by the loss of methanol from m/z 105 and 163 were difficult to interpret, possibly due to their low content, the several methyl groups, and stable isotope ion peaks. The 2 mass differences in the three ion peaks between H_2^{16}O and H_2^{18}O reactions correspond to those between ^{16}O and ^{18}O . No 2 mass different ion peak was observed except m/z 105, 163, and 193, and no m/z 135 ($133 + 2$) ion was observed (Fig. 5). These fragmentation ion peaks may suggest that ^{18}O existed on the C5 atom of Δ GlcA but not on the C1 atom after hydrolysis with ^{18}O water.

In conclusion, the structure of UGL in a complex with unsaturated chondroitin disaccharide provides the basis for substrate binding and insights into the novel catalytic reaction mechanism for glycoside hydrolase. Although the chemical mechanism of acid-catalyzed hydration of the vinyl ether group and the similar mechanism in glycosidases are well characterized, the reaction mechanism involving the hydration of the vinyl ether group triggering the hydrolysis of the glycosidic bond is novel and intriguing. Almost all glycoside-related enzymes (hydrolases and

phosphorylase) with a α_6/α_6 -barrel (six-hairpin glycosidase superfamily of the SCOP database; <http://scop.mrc-lmb.cam.ac.uk/scop>), such as glucoamylases, endoglucanase CelD, endoglucanase CelA, endo/exocellulase E4, and maltose phosphorylase, are “inverting glycosidases.” The CAZy database divides these enzymes into two clans, i.e., clan GH-L (glucoamylase and maltose phosphorylase) and GH-M (cellulase), based on the active site architecture. UGL is therefore the first α_6/α_6 -barrel enzyme with hydration of the vinyl ether group and hence may be classified into a new clan.

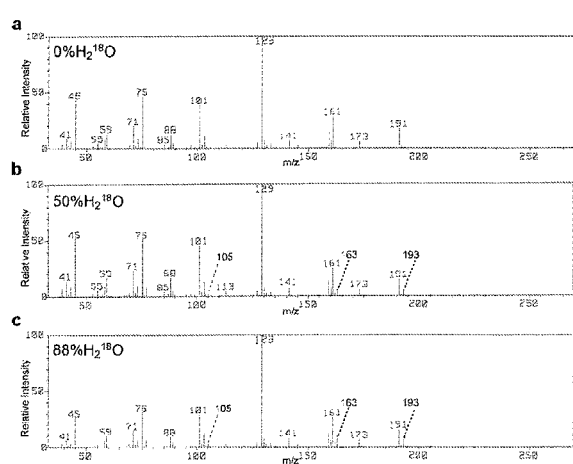
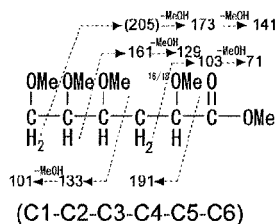


Fig. 5. Mass spectra of Δ GlcA after reduction and methylation. Hydrolysis products incubated with Δ GlcA-GalNAc and UGL at 30°C overnight were reduced, methylated, and subjected to GC/MS analysis. (a) UGL hydrolysis with 0% ^{18}O water. The inset shows the fragmentation pattern of the reduced and methylated Δ GlcA (Chai *et al.*, 1995). (b) UGL hydrolysis with 50% ^{18}O water. (c) Hydrolysis with 88% ^{18}O water.



References

- Aoyama, H., Tokunaga, M., Hiraiwa, S., Shirogane, Y., Obora, Y., and Tsuji, Y. (2004) *Org. Chem.* **6**, 509-512
- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905-921
- Chai, C., Rosankiewicz, J. R., and Lawson, A. M. (1995) *Carbohydr. Res.* **269**, 111-124
- Chou, W. K., Hinderlich, S., Reutter, W., and Tanner, M. E. (2003) *J. Am. Chem. Soc.* **125**, 2455-2461
- Ciucanu, L., and Kerek, F. (1984) *Carbohydr. Res.* **131**, 209-217
- Convington, A. K., Paabo, M., Robinson, R. A., and Bates, R. G. (1968) *Anal. Chem.* **40**, 700-706
- Cremer, D., and Pople, J. A. (1975) *J. Am. Chem. Soc.* **97**, 1354-1358
- Davies, G. J., Wilson, K. S., and Henrissat, B. (1997) *Biochem. J.* **321**, 557-559
- Davies, G., and Henrissat, B. (1995) *Structure* **15**, 853-859
- Fife, T. H., and Brod, L. H. (1970) *J. Am. Chem. Soc.* **92**, 1681-1684

- Hashimoto, W., Kobayashi, E., Nankai, H., Sato, N., Miya, T., Kawai, S., and Murata, K. (1999) *Arch. Biochem. Biophys.* **368**, 367-374
- Hashimoto, W., Nankai, H., Mikami, B., and Murata, K. (2003) *J. Biol. Chem.* **278**, 7663-7673
- Hooft, R. W., Vriend, G., Sander, C., and Abola, E. E. (1996) *Nature* **381**, 272
- Jedrzejewski, M. J., Mello, L. V., de Groot, B. L., and Li, S. (2002) *J. Biol. Chem.* **277**, 28287-28297
- Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24**, 946-950
- Kresge, A. J. (1987) *Acc. Chem. Res.* **20**, 364-370
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26**, 283-291
- Laughton, P. M., and Robertson, R. E. (1969) In *Solvent isotope effects for equilibria and reactions in solute-solvent interactions*, Coetzee JG, Richie CD (eds) pp 407-412, Marcel Dekker, New York
- Lehmann, J., and Schröter, E. (1972) *Carbohydr. Res.* **23**, 359-368
- Ludwigs, U., Elgavish, A., Esko, J. D., Meezan, E., and Rodén, L. (1987) *Biochem. J.* **245**, 795-804
- Lunin, V. V., Li, Y., Linhardt, R. J., Miyazono, H., Kyogashima, M., Kaneko, T., Bell, A. W., and Cygler, M. (2004) *J. Mol. Biol.* **337**, 367-386
- Maruyama, Y., Hashimoto, W., Mikami, B., and Murata, K. (2005) *J. Mol. Biol.* **350**, 974-986
- Merritt, E. A., and Murphy, M. E. P. (1994) *Acta Crystallogr. D Biol. Crystallogr.* **50**, 869-873
- Morgan, P. M., Sala, R. F., and Tanner, M. E. (1997) *J. Am. Chem. Soc.* **119**, 10269-10277
- Nicholls, A., Sharp, K., and Honig, B. (1991) *Proteins Struct. Funct. Genet.* **11**, 281-296
- Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276**, 307-326
- Oyama, K., and Tidwell, T. T. (1976) *J. Am. Chem. Soc.* **98**, 947-951
- Parsons, J. F., Calabrese, K., Eisenstein, E., and Ladner, J. E. (2003) *Biochemistry* **42**, 5684-5693
- Rye, C. S., and Withers, S. G. (2000) *Curr. Opin. Chem. Biol.* **4**, 573-580
- Schlesselman, P., Fritz, H., Lehmann, J., Uchiyama, T., Brewer, C. F., and Hehre, E. J. (1982) *Biochemistry* **21**, 6606-6614
- Speck, A. L. (2001) *PLATON, a multipurpose crystallographic tool*, Utrecht University, The Netherlands
- Weiser, W., Lehmann, J., Chiba, S., Matsui, H., Brewer, C. F., and Hehre, E. J. (1988) *Biochemistry* **27**, 2294-2300
- Zsiska, M., and Meyer, B. (1993) *Carbohydr. Res.* **243**, 225-258

Section 2

Substrate Recognition of Unsaturated Glucuronyl Hydrolase

A novel catalytic mechanism specific for UGL was proposed based on the crystal structure complexed with chondroitin disaccharide (D88N/ Δ GlcA-GalNAc, PDB accession no. 2AHG) in Section 1, Chapter III. The substrate recognition by UGL, however, remains to be clarified. This section deals with the identification of amino acid residues responsible for binding substrates through determination of crystal structures of UGL in complexes with two different substrates (unsaturated hyaluronan disaccharide and gellan tetrasaccharide). Based on the results obtained, the substrate recognition by UGL was clarified.

Materials and Methods

Enzyme assay

UGL reactions were conducted at 30°C as follows. The reaction mixture (0.5 ml) consisted of 50 mM sodium phosphate (pH 6.5), 10-500 μ M of the substrate (Δ GlcA-Glc-Rha-Glc), and 10 nM of the enzyme. Enzyme activity was measured by monitoring the decrease in absorbance at 235 nm; this corresponded to the loss of the C=C double bond of the substrate because the released Δ GlcA is converted to α -keto acid through the loss of the double bond (Fig. 1d, Introduction). Enzyme concentration was determined by UV spectrophotometry using the theoretical molar extinction coefficient $\epsilon_{280} = 99,570$ ($M^{-1}cm^{-1}$). The unsaturated gellan tetrasaccharide (Δ GlcA-Glc-Rha-Glc) used as the UGL substrate was prepared. The parameters k_{cat} and K_m were determined by nonlinear fitting to the Michaelis-Menten equation. The enzyme reaction was also performed in the presence of various concentrations (0-100 mM) of an inhibitor (glycine). The type of inhibition was determined by the Dixon plot ($[I]_0$ vs. $[E]_0/V_0$) at two substrate concentrations ($[S]_0 = 56$ and 115 μ M).

Crystallization and X-ray diffraction

UGL and its mutant D88N from *Bacillus* sp. GL1 were overexpressed in *Escherichia coli*, purified, and crystallized by the vapor diffusion method. Orthorhombic crystals of D88N were placed into the substrate solution containing 400 mM Δ GlcA-GlcNAc (Seikagaku Corporation, Tokyo, Japan) or 400 mM Δ GlcA-Glc-Rha-Glc in a mixture of 25% (w/v) polyethylene glycol 10000 and 0.15 M Tris-HCl (pH 7.5); the crystals were incubated for 10 min at 20°C. Subsequently, they were placed in a cold nitrogen gas stream at -173°C. X-ray

diffraction images of the crystals were collected using a MarCCD detector (Mar USA, Illinois, USA) at the BL-41XU station of SPring-8 (Hyogo, Japan) or a Rigaku Jupiter 210 CCD detector (Tokyo, Japan) at the BL-38B1 station of SPring-8 with synchrotron radiation at a wavelength of 1.00 Å. The images were processed using the DENZO and SCALEPACK softwares (Otwinowski *et al.*, 1997) to a resolution of 1.73 to 1.90 Å (Table 1).

Structure determination and refinement

The structures were determined by the molecular replacement method and refined using CNS program ver. 1.1 (Brünger *et al.*, 1998); the UGL structure (PDB accession no. 1VD5) was used as the reference model. Several rounds of positional and *B*-factor refinement were performed. Following this, manual model building was attempted using the TURBO-FRODO program

Table 1. Synchrotron radiation data collection and refinement statistics

	D88N/ Δ GlcA-Glc-Rha-Glc	D88N/ Δ GlcA-GlcNAc	Hex_wild type/apo
Space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 6 ₃ 22
Unit cell parameters (Å)	a = 87.7, b = 93.5, c = 95.4	a = 87.7, b = 94.9, c = 95.4	a = b = 102.8, c = 223.2
Data collection			
Resolution limit (last shell) ^a (Å)	50.0–1.90 (1.97–1.90)	50.0–1.73 (1.79–1.73)	50.0–1.80 (1.86–1.80)
Measured reflections	284976	429344	646399
Unique reflections	62121 (6030)	79850 (6526)	63925 (6265)
Redundancy	4.6 (4.0)	5.4 (4.2)	10.1 (6.2)
Completeness ($\sum I > \langle I \rangle$) (%)	99.2 (97.6)	95.3 (78.9)	98.2 (98.4)
<i>R</i> _{merge} (%) ^b	5.7 (22.4)	7.1 (33.3)	8.7 (30.8)
Refinement			
Final model	377 × 2 residues, 692 water, Δ GlcA-Glc-Rha-Glc	377 × 2 residues, 789 water, 2 Δ GlcA-GlcNAc	377 residues, 396 water, 9 MPD
Resolution limit (Å)	15.0–1.90 (1.97–1.90)	15.0–1.73 (1.79–1.73)	15.0–1.80 (1.86–1.80)
Used reflections	60367 (5147)	79669 (6567)	64105 (6291)
Completeness ($\sum F_o > \langle F_o \rangle$) (%)	99.0 (83.3)	95.4 (79.6)	98.4 (98.5)
Average <i>B</i> -factor (Å ²)	22.6	19.9	29.1
<i>R</i> -factor (%) ^c	19.8 (24.4)	17.9 (26.5)	16.7 (24.6)
<i>R</i> _{free} (%) ^d	25.4 (31.2)	21.1 (29.3)	18.8 (28.8)
Root-mean-square deviations			
Bond (Å)	0.005	0.005	0.005
Angle (deg)	1.20	1.20	1.18

^aData in highest resolution shells are given in parentheses.

^b $R_{\text{merge}} = \sum |I_i - \langle I \rangle| / \sum I_i \times 100$, where I_i is the intensity of individual reflection and $\langle I \rangle$ is the mean intensity of all reflections.

^c $R\text{-factor} = \sum |F_o - F_c| / \sum F_o \times 100$, where F_o is the observed structure factor and F_c is the calculated structure factor.

^d R_{free} was calculated from 10% of reflections that were randomly selected, as defined by the CNS.

(AFMB-CNRS, Marseille, France) to improve the model by increasing the resolution of the data from 1.73 to 1.90 Å with the CNS program (Table 1). The stereo quality of the model was assessed using the PROCHECK (Laskowski *et al.*, 1993) and WHAT-CHECK (Hooft *et al.*, 1996) programs. Structural alignment was carried out by superimposing with the aid of a fitting program in TURBO-FRODO. Ribbon plots and surface models were prepared using the MOLSCRIPT (Kraulis, 1991), Raster3D (Merrit & Murphy, 1994), and GRASP (Nicholls *et al.*, 1991) programs.

Results

Inhibition by glycine

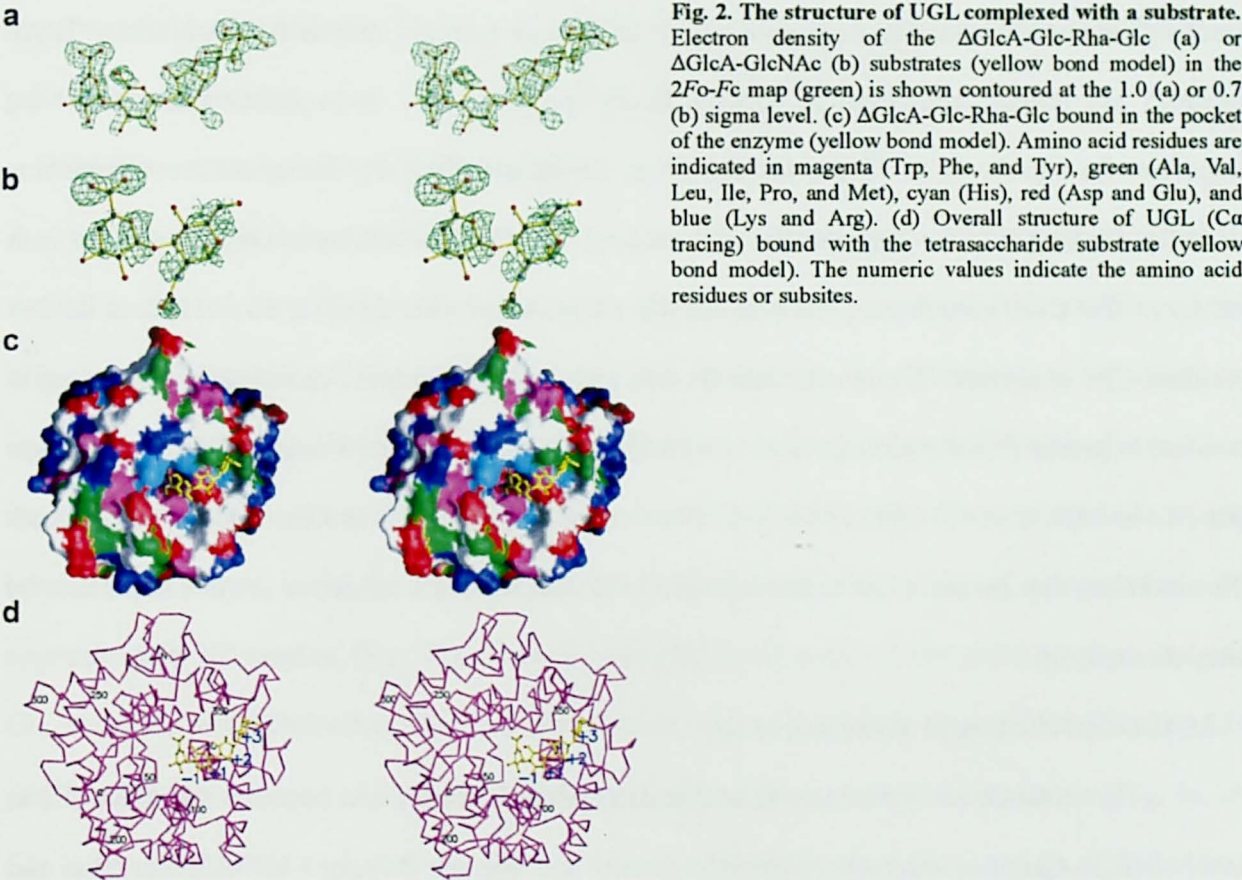
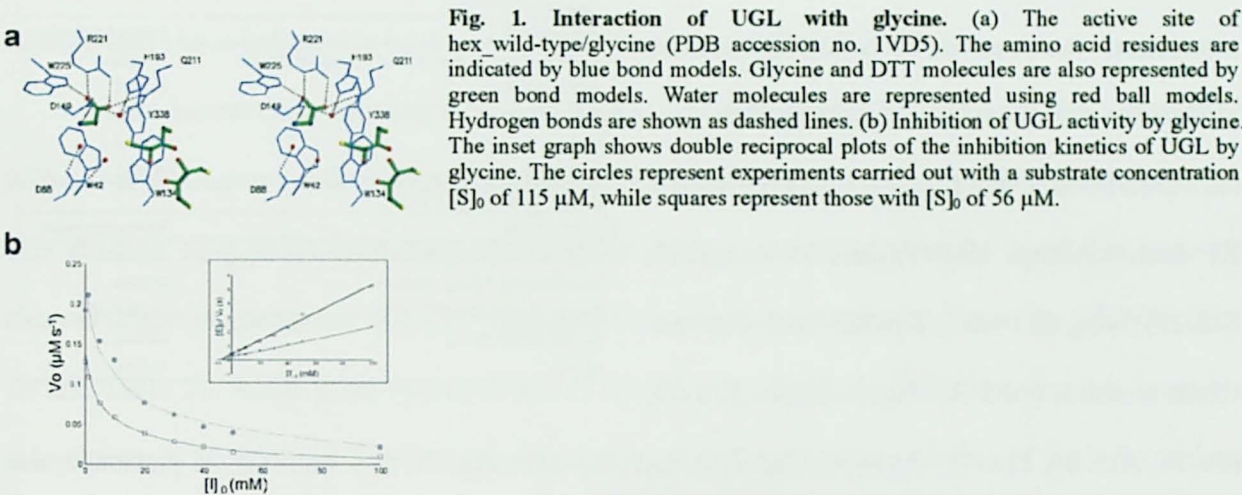
The hex_wild-type/glycine structure, which was the first UGL structure that was determined (PDB accession no. 1VD5) in Chapter I, contains one glycine molecule derived from the crystallization solution. Based on the enzyme-substrate complex structures that are described later, the location of this glycine was found to correspond to that of the carboxyl group of ΔGlcA at the nonreducing terminus of the substrate (Fig. 1a). Next, the effect of glycine on UGL activity was examined. Enzyme activity was inhibited by glycine in a dose-dependent manner (Fig. 1b). The results obtained with the double reciprocal plot revealed that glycine competes with the substrate to bind at subsite -1, which is located at the bottom of the active site, and has a K_i value of 6.5 mM (Fig. 1b, inset graph).

Structure determination

Two orthorhombic crystal complex structures (D88N/ΔGlcA-Glc-Rha-Glc and D88N/ΔGlcA-GlcNAc) were determined (Fig. 2a, b). Results of data collection and refined models are summarized in Table 1.

The hexagonal and orthorhombic crystals are similar in their overall structures, with the exception of their active sites. The active site of the enzyme is situated at the bottom of the large pocket that is surrounded by aromatic amino acid residues; the pocket is formed by the long loops

and the inner helices of the barrel (Fig. 2c, d).



$\Delta\text{GlcA-Glc-Rha-Glc}$ bound at the active site

The unsaturated gellan tetrasaccharide ($\Delta\text{GlcA-Glc-Rha-Glc}$) was bound along the wall leading to the active site and fitted into a concavity of the wall (Figs. 2b and 3a). UGL strongly recognizes ΔGlcA at subsite -1 rather than the saccharides at subsites +1 to +3. Contrary to most

glycosidase-substrate complexes, no direct interaction was observed between UGL and the glycosidic bond to be cleaved. Δ GlcA was stabilized by several tight hydrogen bonds (<3.5 Å) from residues His-87, Asn-88, Asp-149, Gln-211, Arg-221 and Trp-225 and van der Waals contacts (<4.5 Å) from residues Trp-42,

Table 2. Interactions between UGL and substrate

Hydrogen bonds (<3.5 Å)	
Protein residues	Subsites (saccharide)
His-87, Asn(Asp)-88, Asp-149, Gln-211, Arg-221, Trp-225	-1 (Δ GlcA)
Tyr-338, His-339	+1 (Glc)
Trp-134, Gln-211	+1 (GlcNAc)
His-339	+2 (Rha)
Ser-345	+3 (Glc)
Hydrophobic interactions (<4.5 Å)	
Protein residues	Subsites (saccharide)
Trp-42, His-87, Asn(Asp)-88, Phe-91, Asp-149, Trp-219, Arg-221	-1 (Δ GlcA)
Trp-42, Gln-211, Tyr-338, His-339	+1 (Glc)
Trp-134, Tyr-338, His-339	+1 (GlcNAc)
Tyr-338, His-339, Ile-344	+2 (Rha)
Ile-344	+3 (Glc)

His-87, Asn-88, Phe-91, Asp-149, Trp-219, and Arg-221 (Fig. 3a and Table 2). In particular, the stacking interaction with Trp-42 was observed to play an essential role in Δ GlcA binding. These residues are strictly conserved in GH-88 family proteins. The hydrophobic interactions are important for subsite +1 to +3 saccharides (Figs. 2c, 3a and Table 2). Glc at subsite +1 interacts with Trp-42, Gln-211, Tyr-338, and His-339 residues. The His-339 residue in UGL is replaced with serine in the UGL homologues. Rha at subsite +2 interacts with Tyr-338, His-339, and Ile-344 residues. Glc at subsite +3 interacts with Ile-344 and Ser-345 residues. Variations are observed in residues responsible for interactions of saccharides at subsites +1 to +3 among UGL homologues (Ile-344 in UGL is His or Lys in other GH-88 proteins; Ser-345 is Ala or Gly; Fig. 2, Introduction). This indicates that the strict interaction between UGL and Δ GlcA at subsite -1 is more important for enzyme catalysis.

Δ GlcA-GlcNAc bound at the active site

The unsaturated hyaluronan disaccharide (Δ GlcA-GlcNAc) was bound at the active site in D88N/ Δ GlcA-GlcNAc, although its electron density and average B -factor (46.3 Å²) were thin and large, respectively (Fig. 2b). The side-chain of Asp-149 was modeled with two alternate conformations. These features differed from those for Δ GlcA-GalNAc in D88N/ Δ GlcA-GalNAc (average B -factor, 30.8 Å²). However, the molecular weight of Δ GlcA-GlcNAc was the same as that of Δ GlcA-GalNAc since both are isomeric at the C4 hydroxyl group. The substrate concentration of

the complex solution is much higher than its K_m value (approximately 2000 fold). Thus, $\Delta\text{GlcA-GlcNAc}$ is believed to be disordered at the active site (Figs. 2b and 3 b).

The interactions between ΔGlcA and UGL are common to the structures that complexed with either unsaturated chondroitin disaccharide, hyaluronan disaccharide, or gellan tetrasaccharide. The absence of a direct interaction between the enzyme and the glycosidic bond that is to be cleaved is also characteristic of UGL. Hydrophobic interactions are also important for GlcNAc (Fig. 3b and Table 2). While some hydrogen bonds are involved in binding GlcNAc, the two aromatic side-chains of Tyr-338 and Trp-134 are observed to sandwich the pyranose ring of GlcNAc and are important for accommodating GlcNAc at subsite +1, as in the case of unsaturated chondroitin or gellan oligosaccharide.

Wild-type crystal structure without glycine

The crystal structure of the wild-type enzyme without glycine (hex_wild-type/apo) is similar to that of other UGLs; these structures differ only with respect to the position of the side-chains of Asp-149 and Trp-134 at the active site, the position of which changes depending on ligand binding (Fig. 4). The side-chain of Asp-149 forms a hydrogen bond with that of Asp-88 in the hex_wild-type/apo, while in the complex (e.g., D88N/ $\Delta\text{GlcA-Glc-Rha-Glc}$), it forms a hydrogen bond with a water molecule (Wat-1) and is directed to the C4 atom of the substrate by approximately 70° rotation (Fig. 4b). Although two MPD molecules in hex_wild-type/apo and one Glc molecule in D88N/ $\Delta\text{GlcA-Glc-Rha-Glc}$ show similar hydrophobic interactions with Trp-134 and Tyr-338, the direction of the side-chain of Trp-134 is different across the structures (Fig. 4a, c). Glc is restrained at the α -glycoside bond, and these results in the two pyranose rings of ΔGlcA and Glc almost being parallel. The MPD molecule is smaller than saccharide. Thus, the two MPD molecules could bind to two sites in the space between Trp-134 and Tyr-338, accompanying the conformational change of the side-chain of Trp-134. In other words, the side-chain of Trp-134 can suitably reorient by rotating approximately 120°, depending on the type of ligand attached (Fig. 4c).

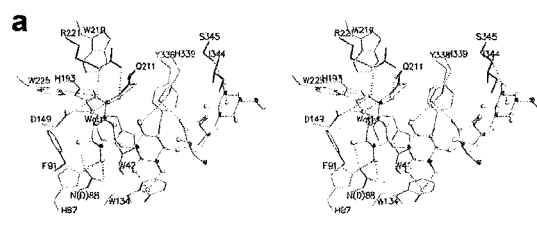


Fig. 3. Structure of UGL bound to the substrate at the active site. (a) D88N/ΔGlcA-Glc-Rha-Glc and (b) D88N/ΔGlcA-GlcNAc. Residues interacting with the substrate are represented by the bond models. Side-chains are indicated in red (Asp and Asn), blue (Arg), purple (Trp, Tyr, Phe), cyan (His), and green (Val, Ser, Ile, and Gln). The substrate is denoted by a yellow bond model. Water molecules are represented as light blue balls. Hydrogen bonds are shown as dashed lines. (c) Structural comparison of D88N/ΔGlcA-Glc-Rha-Glc (green), D88N/ΔGlcA-GlcNAc (magenta), and D88N/ΔGlcA-GalNAc (blue) in the active site.

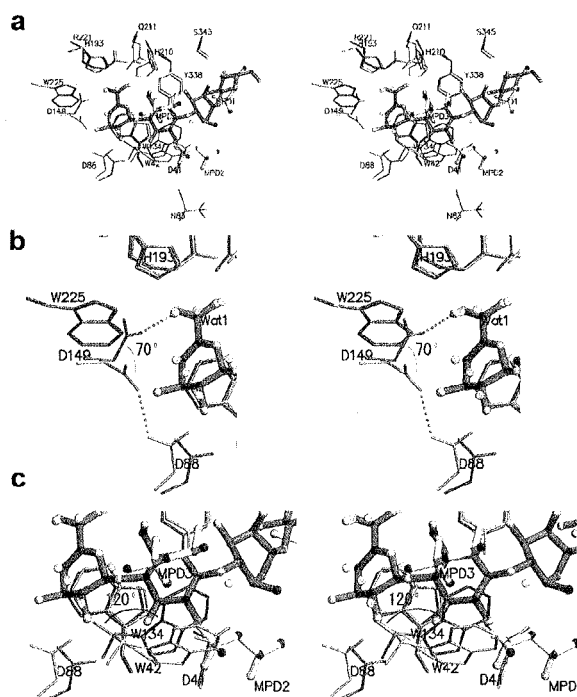
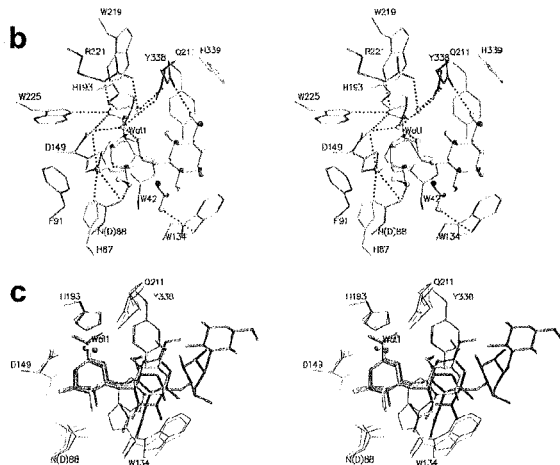


Fig. 4. Active site structure. (a) Structural comparisons of hex wild-type/apo (red) and D88N/ΔGlcA-Glc-Rha-Glc (blue). MPD molecules of hex wild-type/apo are denoted by yellow bond models, while ΔGlcA-Glc-Rha-Glc is denoted by the green bond model. (b) The surroundings of Asp-149 side-chain. The side-chain rotates by 70° to facilitate substrate binding. (c) The surroundings of the Trp-134 side-chain. The side-chain rotates by 120° to facilitate the substrate binding.

Discussion

In this section, the crystal structures of UGL complexed with various substrates clarify the substrate recognition mechanism of the enzyme. Δ GlcA is specifically recognized at subsite -1 through the formation of several hydrogen bonds and stacking interactions in which highly conserved residues of the GH-88 family participate (Fig. 3 and Table 2). In particular, the carboxyl group of Δ GlcA interacts with the enzyme through the formation of hydrogen bonds, and it is stabilized by the positive end of the inner α -helix dipole of the α_6/α_6 -barrel. This observation has been supported by an inhibition experiment performed using glycine. The glycine molecule, which also contains a carboxyl group, could bind to the active site of UGL and competitively inhibits the enzyme (Fig. 1).

The unsaturated gellan tetrasaccharide (Δ GlcA-Glc-Rha-Glc) was bound from the bottom of the pocket to the wall that was completely covered with aromatic or hydrophobic amino acid residues (Fig. 2c). Although several amino acid residues at subsite +3 are not conserved in GH-88 family proteins and some hydrogen bonds are involved in binding the substrate, hydrophobic interactions are important for binding the saccharides at subsites +1 to +3 (Figs. 2c, 3 and Table 2). Subsite +1 is mainly formed in the space between the side-chains of Tyr-338 and Trp-134. The side-chain of Tyr-338 forms hydrogen bonds with Trp-42 and Asp-41. Thus, Tyr-338 is stabilized and does not change its position. However, the side-chain of Trp-134 has high flexibility and can move suitably by rotation in order to bind substrates or ligands (Figs. 3 and 4). This flexibility is believed to provide a wider range of substrate specificity at subsite +1 (e.g., GalNAc, GlcNAc, Man, or Glc). The concavities at subsites +2 and +3 are mainly formed by the aromatic (Tyr-338 and His-339) and hydrophobic (Ile-344) side-chains, which contribute to hydrophobic interactions with the substrate. The two pyranose rings fit into this hydrophobic surface at subsites +2 and +3. In comparison with the strong recognition for Δ GlcA through formation of hydrogen bond networks, the interactions between the enzyme and the saccharides at subsites +2 and +3 are very weak. The

amino acid residues that participate in the interactions at these subsites are not highly conserved in the GH-88 family. It is believed that subsites +2 and +3 are not essential for catalysis.

On the other hand, the catalytic mechanism of UGL has been clarified in which the enzyme triggers the hydration of the vinyl ether group in Δ GlcA (Fig. 4, Section 1, Chapter III). In order to donate a proton to the C4 atom of Δ GlcA, which acts as a general acid catalyst, Asp-149 must be protonated. In the substrate-free structure (hex_wild-type/apo), an unusual contact between Asp-149 and Asp-88 through the formation of a hydrogen bond was observed (Fig. 4b). This hydrogen bond would provide the protonation of Asp-149. It is speculated that Asp-88 may act as a proton donor for Asp-149 and modulate the pKa of Asp-149. This aspartic acid dyad is also observed in cellobiohydrolase Cel6A from *Trichoderma reesei* (Koivula *et al.*, 2002; Wohlfahrt *et al.*, 2003). Asp-88 plays a crucial role not only in controlling the pKa of Asp-149 but also in connecting protons of the hydroxyl groups of Δ GlcA O2 and O3 through the formation of hydrogen bonds. Otherwise, the O3 hydroxyl groups would prevent Asp-149 protonation and also direct it away from C4 and toward O3.

The average *B*-factor of Δ GlcA-GlcNAc was much higher than that of Δ GlcA-GalNAc although Δ GlcA-GlcNAc and Δ GlcA-GalNAc are isomeric at the C4 hydroxyl group. The side-chain of Asp-149 was modeled with two alternate conformations. The two alternate positions in D88N/ Δ GlcA-GlcNAc were similarly located at the positions in the substrate binding or the free structures. The *k*_{cat} value ($0.022 \mu\text{M}^{-1}\text{s}^{-1}$) of Δ GlcA-GlcNAc is much smaller than that of Δ GlcA-GalNAc ($0.075 \mu\text{M}^{-1}\text{s}^{-1}$), although their *K*_m values are comparable. Thus, it was considered that the binding arrangement of Δ GlcA-GalNAc was more favorable for catalysis than that of Δ GlcA-GlcNAc. GlcNAc causes steric hindrance to the active site in comparison with GalNAc. The *N*-acetyl group of GlcNAc also moves and affects the position of Trp-134 (Fig. 3c). As a result, it changes the relationships among Asp-88, Asp-149, and the O2 and O3 hydroxyl groups of Δ GlcA. This movement probably prevents the proper orientation of the Asp-149 side-chain for catalysis and

reduces the *k*_{cat} value. In conclusion, these enzyme complex structures clarify the active site architecture of the four subsites. UGL strongly recognizes ΔGlcA at subsite -1 through the formation of hydrogen bonds and stacking interactions, and prefers GalNAc and Glc rather than GlcNAc as a residue accommodated in subsite +1, due to the steric hindrance.

References

- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905-921
- Hooft, R. W., Vriend, G., Sander, C., and Abola, E. E. (1996) *Nature* **381**, 272
- Koivula, A., Ruohonen, L., Wohlfahrt, G., Reinikainen, T., Teeri, T. T., Piens, K., Claeysens, M., Weber, M., Vasella, A., Becker, D., Sinnott, M. L., Zou, J. Y., Kleywegt, G., Szardenings, M., Stahlberg, J., and Jones, T. A. (2002) *J. Am. Chem. Soc.* **124**, 10015-10024
- Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24**, 946-950
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26**, 283-291
- Merrit, E. A., and Murphy, M. E. P. (1994) *Acta Crystallogr. D Biol. Crystallogr.* **50**, 869-873
- Nicholls, A., Sharp, K., and Honig, B. (1991) *Proteins Struct. Funct. Genet.* **11**, 281-296
- Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276**, 307-326
- Wohlfahrt, G., Pellikka, T., Boer, H., Teeri, T. T., and Koivula, A. (2003) *Biochemistry* **42**, 10095-10103

Section 3

Structure of Unsaturated Galacturonyl Hydrolase Complexed with Substrate

YteR and its homologous protein YesR from *Bacillus subtilis* strain 168 were identified as the unsaturated galacturonyl hydrolase and categorized into a novel glycoside hydrolase family 105 in Chapter II. YteR and YesR act specifically on unsaturated RG produced from RG type-I treated with RG lyases and release Δ GalA from the substrate. This section deals with the clarification of catalytic mechanism and substrate recognition of YteR through the determination of crystal structure of its mutant (D143N) in complex with the intrinsic substrate (Δ GalA-Rha).

Materials and Methods

Mutagenesis

Asp-143 in YteR (or Asp-135 in YesR) was replaced with an asparagine residue by the use of a QuickChange site-directed mutagenesis kit (Stratagene) and mutations were confirmed by DNA sequencing. The plasmid, which is the expression vector for YteR (or YesR), was used as a template. Primers were as follows: YteR/D143N, 5'-CCAAATGTGGCTTAATGGCTTATACATGGG-3' and 5'-CCCATGTATAAGCCATTAAGCCACATTTGG-3' (mutations are indicated by bold letters) (or YesR/D135N, 5'-GAACAGGCATGGGCGAATACGCTGATGATG-3' and 5'-CATCATCAGCGTATTCGCCCCATGCCTGTTC-3'). Mutant proteins were expressed and purified as done for the wild-type described in Chapter II.

Crystallization and data collection

D143N was crystallized by hanging drop vapor diffusion described by Zhang *et al.* (2005). Substrate (Δ GalA-Rha) for YteR was prepared using potato RG-I (Megazyme International Ireland Ltd., Wicklow, Ireland) and RG lyases (YesW and YesX from *B. subtilis* strain 168) as described in Chapter II. A single crystal was soaked in the solution containing 120 mM Δ GalA-Rha, 1.2 M

citrate and 0.1 M Tris-HCl (pH 7.5) for 15 min at 20°C.

The crystal was placed in a cold nitrogen gas stream at -173°C. X-ray diffraction images were collected using a Jupiter210 area detector (Rigaku, Tokyo, Japan) with synchrotron radiation at a wavelength of 1.0 Å at the BL-38B1 station of SPring-8 (Hyogo, Japan). Images were processed to a resolution of 1.9 Å using DENZO and SCALEPACK software (Table 1) (Otwinowski *et al.*, 1997).

Structure determination and analysis

The crystal structure was determined by molecular replacement using the CNS ver. 1.1 program (Brünger *et al.*, 1998) with the reported YteR wild-type/apo structure as the reference model. Coordinates of UGL mutant D88N/ΔGlcA-GalNAc complex (2AHG) and YteR wild-type/apo (1NC5) were obtained from the RCSB Protein Data Bank (<http://pd-beta.rcsb.org>). Several rounds of positional and *B*-factor refinement followed by manual model building using the TURBO-FRODO program (AFMB-CNRS, Marseille, France) were done to improve the model using the CNS program to a resolution

Table 1. Data collection and refinement statistics

Data collection	
Resolution (last shell) ^a (Å)	50.0-1.90 (1.97-1.90)
Measured reflections	270713
Unique reflections	35001 (3106)
<i>I</i> / <i>σ</i> (<i>I</i>)	29.9 (1.5)
Redundancy	7.7 (3.6)
Completeness (%)	97.7 (88.8)
<i>R</i> _{merge} (%) ^b	7.0 (26.1)
Refinement	
Final model	363 residues (11-373), 287 water, and ΔGalA-Rha molecules
Average <i>B</i> -factor (Å ²)	
Protein	34.2
Water	43.5
ΔGalA	25.5
Rha	34.4
<i>R</i> -factor (%) ^c	18.9 (32.1)
<i>R</i> _{free} (%) ^d	23.3 (33.8)
Root mean square deviations	
Bond (Å)	0.006
Angle (deg)	1.25
Ramachandran plot (%)	
Most favored regions	96.0
Additional allowed regions	4.0
Disallowed	0

^a Data on highest resolution shells is given in parentheses.

^b $R_{\text{merge}} = \sum |I_i - \langle I \rangle| / \sum I_i \times 100$, where I_i is the intensity of individual reflection and $\langle I \rangle$ the mean intensity of all reflections.

^c $R\text{-factor} = \sum |F_o - F_c| / \sum F_o \times 100$, where F_o is the observed structure factor and F_c the calculated structure factor.

^d R_{free} was calculated from randomly chosen 5% reflections.

of 1.9 Å (Table 1). The stereoquality of the model was assessed using the PROCHECK (Laskowski *et al.*, 1993) and WHAT-CHECK (Hoofst *et al.*, 1996) programs. Structural alignment was done by superimposition using a fitting program in the TURBO-FRODO program. Ribbon plots were prepared using the MOLSCRIPT (Kraulis, 1991) and Raster3D (Merritt & Murphy, 1994) programs.

Electrostatic surface potential was calculated using the GRASP (Nicholls *et al.*, 1991) program. The full.crg was used for charge assignment, where the histidine residue has a positive charge. Carbohydrate conformations (puckering parameters) as defined by Cremer and Pople (1975) were analyzed using the PLATON program (Speck, 2001).

Results and Discussion

Structure determination

YteR is a monomeric enzyme with a molecular mass of 43 kDa (373 amino acid residues). Based on structural comparison between YteR and UGL, Asp-143 seems to be a candidate for the catalytic residue. In fact, the mutant D143N of YteR substituting Asn-143 for Asp-143 was inactive as described below. This inactive mutant D143N was crystallized based on crystallization described by Zhang *et al.* (2005) in order to obtain enzyme-substrate complex. Δ GalA-Rha is a disaccharide and good substrate for YteR. The crystal was soaked in this substrate solution. Datasets of the crystal were obtained at 1.9 Å resolution by synchrotron radiation (Table 1). The complex structure (D143N/ Δ GalA-Rha) was determined at 1.9 Å resolution by molecular replacement (Fig. 1 and Table 1). The refined model consists of 363 amino acid residues (Lys-11 to Lys-373), 287 water molecules, and one Δ GalA-Rha molecule. The first Met-1 residue was excised by *E. coli* aminopeptidase. The entire polypeptide chain sequence was well traced, and electron densities of the main- and side-chains were generally very well defined in the $2F_o - F_c$ map, except for N-terminal amino acid 9 residues (Gly-2 to Val-10) whose electron densities were too low to be identified. All nonglycine residues are localized within most favored or additional allowed regions of the Ramachandran plot, as defined by the PROCHECK program (Table 1). The Δ GalA-Rha model also fitted well in the map (Fig. 2a), but the average B -factor of Rha (34.4 Å²) was slightly higher than that of Δ GalA (25.5 Å²) (Table 1). The difference in B -factors is related to the fewer

interactions between Rha and the enzyme, as discussed later. The enzyme molecule has a α_6/α_6 -barrel structure (Fig. 1a). The α_6/α_6 -barrel structure consists of six outer helices running in roughly the same direction and six inner helices oriented in the opposite direction. These helices are connected in a nearest-neighbor and an up-and-down pattern by short and long loops. One side of the α_6/α_6 -barrel structure consists of long loops containing short β -sheets, and contributes to the formation of a pocket.

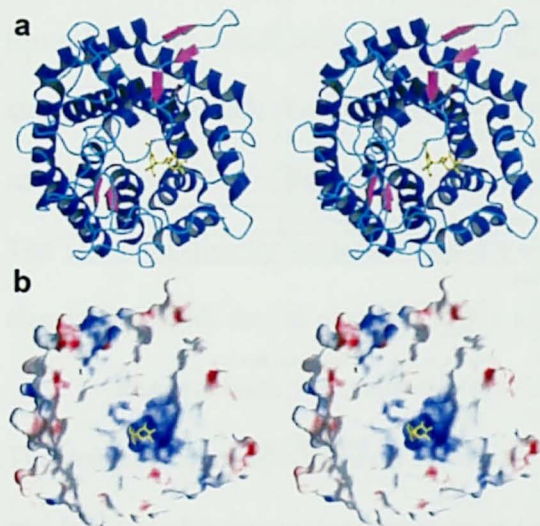


Fig. 1. Structure of YteR complexed with substrate. (a) Overall structure of YteR bound with Δ GalA-Rha. The α_6/α_6 -barrel is shown in blue (helix), magenta (strand), and cyan (loop). The yellow bond model indicates the substrate (Δ GalA-Rha) (oxygen atom, red; carbon atom, black). (b) The electrostatic potential surface of YteR bound with substrate. The structure is represented as a white molecular surface model. The electrostatic potential surface is drawn in the range from $-40k_B T$ (red) to $+40k_B T$ (blue), where k_B is Boltzmann's constant and T is absolute temperature (K). The substrate is also represented by the yellow bond model.

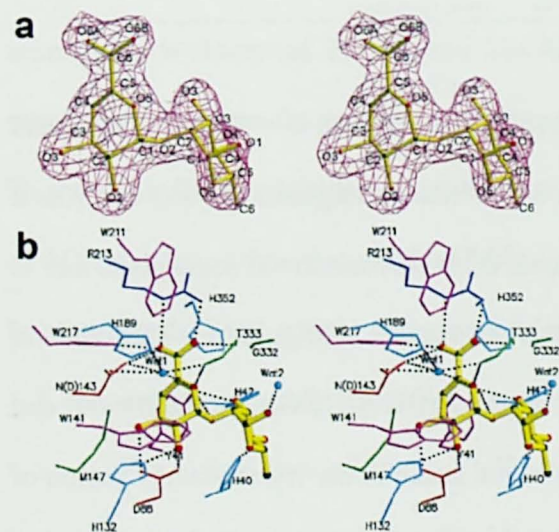


Fig. 2. The substrate bound in the active site. (a) Electron density of Δ GalA-Rha (substrate) in the omit ($F_o - F_c$) map calculated without the substrate and contoured at 3.0σ (magenta). Characters indicate atom names of the substrate. (b) Residues interacting with the substrate are represented by bond models. Residues are indicated in red (Asp and Asn), blue (Arg), purple (Trp, Tyr, and Phe), cyan (His), and green (Thr, Gly, and Met). The substrate is denoted by a yellow bond model (oxygen atom, red; carbon atom, black). Water molecules (Wat-1 and Wat-2) are represented as a cyan ball. Hydrogen bonds are shown as dashed lines.

YteR-ΔGalA-Rha complex

The disaccharide ΔGalA-Rha (substrate) was bound at the bottom of the pocket (Figs. 1 and 2), indicating that it occupied two subsites, -1 and +1, at the active site. The active pocket is formed by the positive end of the inner α -helix of the α_6/α_6 -barrel and long loops. It has a shape with a depth and width of ~ 10 Å and is surrounded by the positive potential derived from the positive side-chains of His-40, His-42, His-132, His-189, His-332, and Arg-213 residues and the inner helix dipole (Figs. 1b and 2b). ΔGalA has a carboxyl group and is a negatively charged saccharide, although Rha is a neutral saccharide. The positive charged pocket is suitable for concentrating the negative substrate. The

puckering parameters of bound ΔGalA-Rha were $Q = 0.48$ Å, $\Theta = 50^\circ$, and $\Phi = 92^\circ$ for ΔGalA, and $Q = 0.59$ Å, $\Theta = 3^\circ$, and $\Phi = 96^\circ$ for Rha. ΔGalA has a half-chair configuration (2H_1) and Rha a stable chair configuration (1C_4). The C3, C4, and C5 atoms of ΔGalA are on the same plane due to the C4=C5 double bond restriction. ΔGalA therefore has an unusual configuration. The dihedral angles of the β -1,2-linkage between ΔGalA and Rha are φ ($O5_{\Delta GalA}-C1_{\Delta GalA}-O2_{Rha}-C2_{Rha}$) = 69° and ψ ($C1_{\Delta GalA}-O2_{Rha}-C2_{Rha}-C1_{Rha}$) = 78° . These torsion angles are stable in the lowest energy region of the isoenergy map (Mazeau & Pérez, 1998).

ΔGalA-Rha interacts with residues, His-40, Tyr-41, His-42, Asp-88, His-132, Trp-141, Asn(Asp)-143, Met-147, Arg-213, Trp-217, Trp-211, Gly-332, Thr-333, and His-352 (Fig. 2b and Table 2). All these residues, excepting for His-40, Tyr-41, Asp-88, and His-352 affecting subsite +1,

Table 2. Interaction between D143N and ΔGalA-Rha at subsite -1 and +1

Hydrogen bonds (< 3.3 Å)			
Atom	Paired atom in D143N		Distance (Å)
ΔGlcA			
O2	Asp-88	O ^{B2}	2.5
O2	His-132	N ^{B2}	2.7
O3	Asp-88	O ^{B1}	2.8
O5	Thr-333	O	2.8
O5	Wat-1		3.1
O6A	Asn-143	N ^{B2}	3.3
O6A	Arg-213	N ^B	3.0
O6A	Trp-217	N ^{B1}	3.0
O6B	Arg-213	N ^{B2}	2.8
O6B	Thr-333	N	2.9
Rha			
O1	Wat-2		2.6
O2	Tyr-41	O ^B	3.3
O3	Tyr-41	O ^B	2.6
O4	His-40	N ^{B2}	2.9
van der Waals contact (C-C distance < 4.5 Å)			
Atom	Paired residue in D143N		
ΔGlcA			
C1	Trp-141, Thr-333		
C2	Asp-88, His-132, Trp-141		
C3	Tyr-41, Asp-88, His-132, Asn-143, Met-147		
C4	Tyr-41, Asn-143, Met-147		
C5	Tyr-41		
C6	Trp-211, Arg-213, Gly-332, His-352		
Rha			
C2	Gly-332		
C3	Tyr-41, His-42, Gly-332		
C4	His-40, Tyr-41, His-42		

are almost conserved in YteR and its homologues. The aromatic side-chain of Tyr-41 shows a stacking interaction with the sugar ring of Δ GalA and thus plays an essential role in binding Δ GalA, although Tyr-41 is Trp in the other YteR homologues. The side-chain of Trp-217 forms a hydrogen bond to the carboxyl group of Δ GalA. Other aromatic side-chains (Trp-211 and Trp-141) partially bind to Δ GalA through hydrophobic interactions. The positively charged side-chain of Arg-213 faces the carboxyl group of Δ GalA and forms a salt-bridge with the group. The carboxyl group also forms hydrogen bonds with Asn(Asp)-143 and Thr-333. Through direct hydrogen bonds, O2 and O3 atoms associate with Asp-88 and His-132 at the bottom surface. Although Asp-88 is thought to be an residue essential in binding Δ GalA with hydrogen bonds, it is Asn in the other YteR homologues. The O5 atom directly interacts with Thr-333. The residues, Asn(Asp)-143, His-189, and Thr-333, also interact with the O5 atom through a water molecule (Wat-1).

Rha interacts with His-40, Tyr-41, His-42, and Gly-332 residues (Fig. 2b and Table 2). These residues, except for Gly-332, are not conserved in YteR and its homologues. Hydrogen bonds and hydrophobic interactions play an essential role in the binding of Rha. Positively charged and aromatic side-chains of His-40 and His-42 face the hydroxyl groups of Rha. The number of interactions between the enzyme and Rha is fewer than that between the enzyme and Δ GalA (Fig. 2b and Table 2). The subsite +1 residues are less conserved in family GH-105. The average *B*-factor of Δ GalA is lower than that of Rha (Table 1). Δ GalA is expected for a more deeply buried than Rha. In other words, YteR strictly recognizes Δ GalA at subsite -1 rather than Rha at subsite +1.

Comparison with substrate-free structure

The ligand-free wild-type YteR structure (wild-type/apo) was determined by Zhang *et al.* in a structural genomics project at the Midwest Center (2005). The overall structures (wild-type/apo and D143N/ Δ GalA-Rha) were essentially identical. No overall conformational change was observed in the binding of Δ GalA-Rha. The root mean square distance was 0.11 Å for the superimposition of all residues (363 C α atoms). At the active site, superimposition (Fig. 3a) shows a

local change in which the side-chain of Tyr-41 shifts about 1 Å closer to glycosidic oxygen (O2 atom of Rha) and forms hydrogen bonds with the glycosidic oxygen and O3 atom of Rha (Table 2). Residues, His-40, His-42, Lys-133, and Trp-217, move slightly, about 0.4-0.7 Å, to the substrate and interact with it. Arg-213 moves about 0.4 Å away from the carboxyl group of ΔGalA and forms hydrogen bonds with the group. The side-chains of Trp-211 and His-352 move about 2.5 Å to the substrate. Based on the local change of His-352, the main-chain (Leu-351 to Gly-353) near the positive end of the 12th α-helix of the α₆/α₆-barrel shifts slightly, about 0.5-1.3 Å. Seven water molecules are excluded from the active site, while specific hydrogen bonds among polar groups and hydrophobic interactions among nonpolar groups are formed between the enzyme and the substrate as described above.

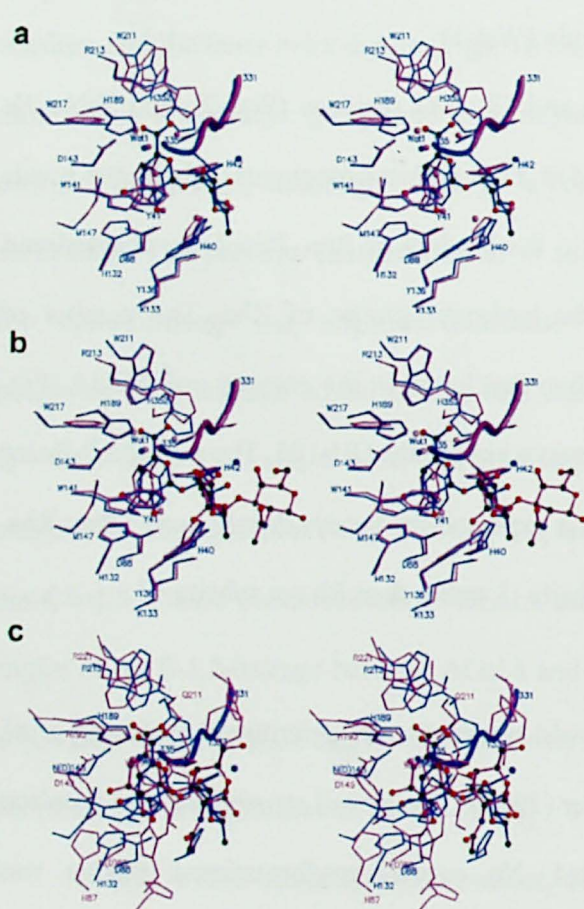


Fig. 3. Structures of ligand-free and bound YteRs. (a) Comparison of wild-type apo form (INC5; magenta) and mutant holo form (D143N/ΔGalA-Rha; blue) at the active site. The ribbon model represents the protruding loop (amino acid residues, 331 through 335). (b) Comparison of D143N/ΔGalA-Rha (blue) and D143N/ΔGlcA-GalNAc (2D8L; magenta) structures at the active site. (c) Comparison of D143N/ΔGalA-Rha (blue) and UGL (D88N)/ΔGlcA-GalNAc complex (2AHG; magenta) structures at the active site.

Comparison to YteR/ΔGlcA-GalNAc structure

The crystal structure of wild-type YteR complexed with the UGL substrate

(Δ GlcA-GalNAc) was determined in Chapter II. Although YteR cannot act on Δ GlcA-GalNAc, it is able to bind the disaccharide. The disaccharide binds at the gate of the active site of YteR (Fig. 3b). The overall structures (wild-type/ Δ GlcA-GlcNAc and D143N/ Δ GalA-Rha) were essentially identical. The root mean square distance was 0.12 Å for the superimposition of all residues (363 C α atoms). At the active site, local change similar to that described above was observed. The side-chain of Tyr-41 moves about 0.9 Å. Residues, His-40, His-42, Arg-213, Trp-211, and Trp-217, move slightly, about 0.4-0.6 Å. The side-chain of His-352 also moves about 2.5 Å and the main-chain (Leu-351 to Gly-353) shifts about 0.5-1.4 Å.

RG consists of a α -1,4-linkage between Rha and GalA. In Δ GlcA-GalNAc, the linkage is α -1,3 between Δ GlcA and GalNAc. The O3 atom of GalNAc appears to share the same position as the O1 atom of Rha (Fig. 3b) and the α -1,3-linkage of bound Δ GlcA-GalNAc is thought to share the same position as the α -1,4-linkage between Rha and the next GalA. Although some hydrogen bonds with water molecules are present, GalNAc does not interact directly with YteR, is highly flexible, and clings to Δ GlcA in the wild-type/ Δ GlcA-GalNAc structure. Thus, subsite +2 for GalA of the unsaturated RG oligomer (trisaccharide <) is thought to interact partially with YteR and to be in the solvent region.

Comparison to GH-88 family UGL

The three-dimensional structure of YteR significantly resembles that of UGL (root mean square distance = 0.6 Å for the superimposition of 284 C α), although they constitute different glycoside hydrolase families (GH-105 and GH-88) and their amino acid sequence identity is negligible (< 10%). In this section, the structure of YteR complexed with Δ GalA-Rha clarifies the active site architecture, in particular, subsites -1 and +1. Ionizable residues, His-132, Asn(Asp)-143, His-189, and Arg-213, interact with the substrate and are conserved in YteR and its homologues, as described above (Fig. 2b). Residues are also highly conserved between YteR and UGL (Fig. 3c). Asp-143 in YteR corresponds to Asp-149 in UGL, as does His-189 to His-193 and Arg-213 to

Arg-221. The space of His-132 in YteR is occupied by His-87 in UGL.

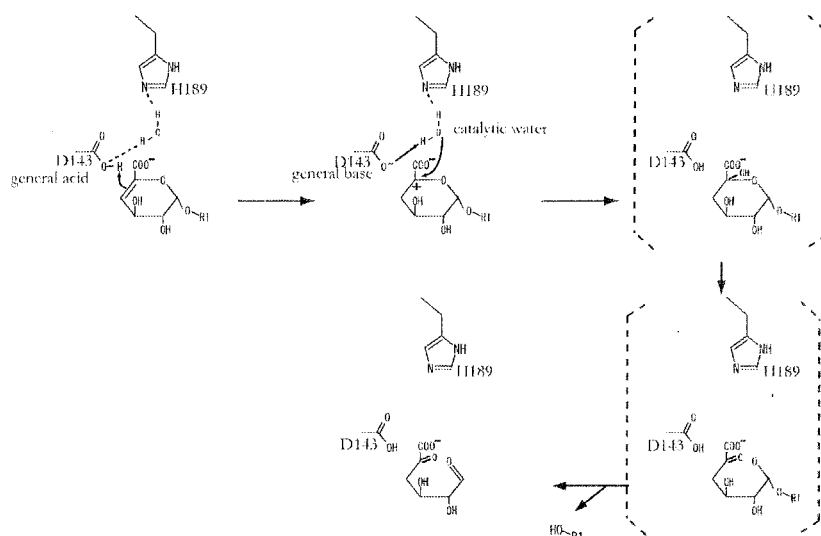


Fig. 4. Proposed catalytic reaction mechanism of YteR. The catalytic reaction proceeds through the water addition reaction of the vinyl ether group, as described in the text. Important residues surrounding the substrate are shown.

Based on the crystal structure of D143N/ Δ GalA-Rha and the proposed catalytic reaction mechanism of UGL, a catalytic reaction mechanism of YteR is also postulated as follows; a water addition reaction (hydration) of the vinyl ether group (C4=C5-O5) and subsequent hydrolysis of the glycoside bond in Δ GalA (Fig. 4) based on the crystal structure of D143N/ Δ GalA-Rha and the proposed catalytic reaction mechanism of UGL. For YteR, Asp-143 can donate a proton to the double bond (C4 atom) of Δ GalA as a general acid catalyst because the location of Asp-143 is suitable for the protonation of the C4 atom of Δ GalA, according to the superimposition with the UGL (D88N)/ Δ GlcA-GalNAc structure (Fig. 3c), and Asp-143 is completely conserved in YteR and its homologues (Fig. 2, Chapter II). In YesR, the residue is predicted to be Asp-135 (Fig. 2, Chapter II). Subsequently, Asp-143 deprotonates the water molecule (Wat-1) as a general base catalyst and the deprotonated water molecule attacks the C5 atom of Δ GalA. The product hemiketal is unstable and readily converts to α -keto acid (hemiacetal). Due to aldehyde-hemiacetal equilibrium, the resultant hemiacetal is finally converted to an aldehyde (4-deoxy-L-threo-5-hexosulose-uronate), leaving saccharide through the cleavage of the glycoside bond (Fig. 4). These consecutive reactions are similar to the oxymercuration of unsaturated hyaluronate disaccharide Δ GlcA-GlcNAc

(Ludwing *et al.*, 1987). The water addition reaction of the vinyl ether group in Δ GlcA and spontaneous conversion of hemiketal to aldehyde, leaving saccharide via α -keto acid (hemiacetal), are also demonstrated in the reaction mixture of Δ GlcA-GlcNAc and mercuric salts. Many glycosidases hydrate unsaturated substrates (D-glycals, C2=C1-O5 or D-glyco-enitols, O5-C1=C). Their catalytic mechanism and the anomeric configuration of products are regulated by their own specificities in hydrolysis, such as “retaining” or “inverting mechanisms” (Lehmann & Schröter, 1972; Hehre *et al.*, 1977, 1980, Brockhaus & Lehmann, 1977; Schlesselmann *et al.*, 1982; Lehmann & Zieger, 1977; Weiser *et al.*, 1988). To investigate candidates, Asp-143 in YteR and Asp-135 in YesR, for catalytic residues, two mutant proteins, YteR/D143N and YesR/D135N, having a substitution of Asn for Asp were assayed. These mutants are completely inactive for the substrate (Δ GalA-Rha). These mutant proteins suggest that Asp-143 in YteR and Asp-135 in YesR are essential for catalysis and support the proposed mechanism that these residues act as the general acid/base.

His-189, which is also completely conserved in YteR and its homologues (Fig. 2, Chapter II), is also located close to Δ GalA and has an ionizable side-chain. It is considered that His-189 probably plays a crucial role in fixing Wat-1 onto a suitable position for catalysis (Fig. 2b).

Although Asp-88 (YteR) is not completely conserved in the GH-105 family, Asp-88 (YteR) and Asp-88 (UGL) play an essential role in typing protons of hydroxyl groups of O2 and, most importantly, O3 in hydrogen bonds (Figs. 2b and 3c). Asp-143 (YteR) and Asp-149 (UGL) as acid catalysts should be protonated. The O3 hydroxyl group would be able to form a hydrogen bond to the side-chain of Asp-143 (YteR) or Asp-149 (UGL). Asp-88 thus prevents the acid catalyst from ionization and also directs them away from O3 and toward the C4 atom of Δ GalA or Δ GlcA.

The optimal pH (pH 4) of the YteR reaction differed from that (pH 6) of the YesR reaction (Fig. 4, Chapter II). Although amino acid residues, whose side-chains are exposed to the surface of the active site, are mostly conserved in the YteR family (Fig. 2b; Fig. 2, Chapter II), two residues,

Tyr-41 (Trp in other enzymes) and Asp-88 (Asn in other enzymes) in YteR are not conserved. Thus, the two residues (Tyr-41 and Asp-88) may be essential to modulating pKa of the Asp-143 carboxyl group to a suitable value and the pH of the enzyme to the optimum.

The five sequential residues, 331 to 335 in YteR, reduce the width of the gate of the active pocket. Due to the protruding loop containing residues from 331 to 335, subsite +1 of YteR is narrower than that of UGL (Fig. 3c). The GalNAc in the UGL complex can not be superimposed on that of Rha in the D143N complex by rotating along the glycoside bond. GalNAc would create steric hindrance with subsite +1 of YteR. The protruding loop of YteR prevents the effective binding of the UGL substrate.

The crystal structure of YteR mutant D143N complexed with the substrate (Δ GalA-Rha), which spans subsites -1 and +1, reveals the active site architecture and the substrate specificity of YteR. YteR strictly recognizes Δ GalA at subsite -1 rather than Rha at subsite +1. Δ GalA is strongly recognized by several tight hydrogen bonds to the carboxyl and hydroxyl groups of Δ GalA and the stacking interaction of Tyr-41. The overall structures (wild-type/apo, wild-type/ Δ GlcA-GlcNAc and D143N/ Δ GalA-Rha) were essentially identical. Although no overall conformational change was observed in the binding of disaccharide, local change in Tyr-41 was found at the active site. The residue shifts to the substrate and plays an essential role in binding the substrate with hydrogen bonds and van der Waals contacts. The general acid/base catalyst and the catalytic water are conserved between YteR (GH-105) and its structural homologous UGL (GH-88) as indicated by comparison to their substrate-enzyme complex structures. The proposed catalytic mechanism is also strongly supported by comparison to the structures of YteR and UGL in complex with their intrinsic substrates. The active pocket of YteR, however, has a specific shape for Δ GalA-Rha. The UGL substrate (Δ GlcA-GalNAc) is repulsed by the protruding loop of YteR, preventing YteR from acting on the UGL substrate.

References

- Brockhaus, M., and Lehmann, J. (1977) *FEBS Lett.* **62**, 154-156
- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905-921
- Cremer, D., and Pople, J. A. (1975) *J. Am. Chem. Soc.* **97**, 1354-1358
- Hehre, E. J., Brewer, C. F., Uchiyama, T., Schlesselmann, P., and Lehmann, J. (1980) *Biochemistry* **19**, 3557-3564
- Hehre, E. J., Genghof, D., Sternlight, H., and Brewer, C. (1977) *Biochemistry* **16**, 1780-1787
- Hooft, R. W., Vriend, G., Sander, C., and Abola, E. E. (1996) *Nature* **381**, 272
- Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24**, 946-950
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26**, 283-291
- Lehmann, J., and Schröter, E. (1972) *Carbohydr. Res.* **23**, 359-368
- Lehmann, J., and Zieger, B. (1977) *Carbohydr. Res.* **58**, 73-78
- Ludwigs, U., Elgavish, A., Esko, J. D., Meezan, E., and Rodén, L. (1987) *Biochem. J.* **245**, 795-804
- Mazeau, K., and Pérez, S. (1998) *Carbohydr. Res.* **311**, 203-217
- Merrit, E. A., and Murphy, M. E. P. (1994) *Acta Crystallogr. D Biol. Crystallogr.* **50**, 869-873
- Nicholls, A., Sharp, K., and Honig, B. (1991) *Proteins Struct. Funct. Genet.* **11** 281-296
- Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276**, 307-326
- Sakamoto, T., Yamada, M., Kawasaki, H., and Sakai, T. (1997) *Eur. J. Biochem.* **245**, 708-714
- Schlesselmann, P., Fritz, H., Lehmann, J., Uchiyama, T., Brewer, C. F., and Hehre, E. J. (1982) *Biochemistry* **21**, 6606-6614
- Speck, A. L. (2001) *PLATON, a multipurpose crystallographic tool*, Utrecht University, The Netherlands
- Weiser, W., Lehmann, J., Chiba, S., Matsui, H., Brewer, C. F., and Hehre, E. J. (1988) *Biochemistry* **27**, 2294-2300
- Zhang, R., Minh, T., Lezondra, L., Korolev, S., Moy, S. F., Collart, F., and Joachimiak, A. (2005) *Proteins* **60**, 561-565

CONCLUSION

1. The X-ray crystallographic structure of UGL from *Bacillus* sp. GL1 was first determined at 1.8 Å resolution. It includes α_6/α_6 -barrel topology which is found in the α/α -toroidal fold in the six-hairpin glycosidase superfamily.
2. YteR and YesR from *Bacillus subtilis* strain 168 were identified as a novel enzyme, “unsaturated galacturonyl hydrolase,” and classified as a new glycoside hydrolase family 105 based on their gene arrangements in genome and enzyme properties. This enzyme acts specifically on Δ GalA-Rha produced from plant cell wall RG type-I treated with RG lyases, releasing Δ GalA from the substrate.
3. The reaction mechanism catalyzed by UGL that triggers the hydration of the vinyl ether group in Δ GlcA was elucidated based on the structural analysis of the UGL-substrate (Δ GlcA-GalNAc) complex and biochemical analysis. This mechanism differs from the generally accepted mechanism in glycosidases.
4. The other structures of UGL-substrate (Δ GlcA-GlcNAc and Δ GlcA-Glc-Rha-Glc) complexes clarified the substrate recognition mechanism of UGL.
5. The structure of YteR in complex with substrate (Δ GalA-Rha) was determined. This structural feature directly contributes to the postulation of the enzyme reaction mechanism.

The coordinates of UGL and YteR structures have been deposited in the RCSB Protein Data Bank (<http://www.rcsb.org/pdb/Welcome.do>).

Hexagonal UGL/Glycine/DTT/MPD (**1VD5**); Hexagonal UGL (**2FUZ**); Orthorhombic UGL (**2D5J**); UGL(D88N)/ Δ GlcA-GalNAc (**2AHG**); UGL(D88N)/ Δ GlcA-GlcNAc (**2FV1**); UGL(D88N)/ Δ GlcA-Glc-Rha-Glc (**2FVO**); YteR(D143N)/ Δ GlcA-GalNAc (**2D8L**)

ACKNOWLEDGEMENTS

The author wishes to express his most sincere gratitude to Dr. Kousaku Murata, Professor of Graduate School of Agriculture, Kyoto University, for his kind guidance and warm encouragement during the course of this study.

The author is greatly indebted to Dr. Bunzo Mikami, Professor of Graduate School of Agriculture, Kyoto University, and Dr. Wataru Hashimoto, Associate Professor of Graduate School of Agriculture, Kyoto University, for their constructive advice and extensive discussion.

The author would like to express his sincere gratitude to Dr. Shigeyuki Kawai, Research Associate of Graduate School of Agriculture, Kyoto University, for his critical comments.

The author is grateful to Dr. Kuniyo Inouye and Dr. Bunzo Mikami, Professors of Graduate School of Agriculture, Kyoto University, for their critical reading of this thesis.

The author is thankful to Dr. Masayuki Yamasaki and Dr. Yukie Maruyama for their technical advices and valuable comments.

The author would like to express his heartfelt thanks to Dr. Shigeru Utsumi, Professor of Graduate School of Agriculture, Kyoto University, for his helpful comments.

Thanks are given to all the members of the Laboratories of Basic and Applied Molecular Biotechnology, of Applied Structural Biology, and of Food Quality Design and Development, Graduate School of Agriculture, Kyoto University, for their helpful collaborations and encouragement.

Takafumi ITOH

LIST OF PUBLICATIONS

Original papers

1. **Takafumi Itoh, Sae Akao, Wataru Hashimoto, Bunzo Mikami, and Kousaku Murata:** Crystal structure of unsaturated glucuronyl hydrolase, responsible for the degradation of glycosaminoglycan, from *Bacillus* sp. GL1 at 1.8 Å resolution.
J. Biol. Chem., **279**(30), 31804-31812 (2004).
2. **Takafumi Itoh, Akihito Ochiai, Bunzo Mikami, Wataru Hashimoto, and Kousaku Murata:** A novel glycoside hydrolase family 105: the structure of family 105 unsaturated rhamnogalacturonyl hydrolase complexed with a disaccharide in comparison with family 88 enzyme complexed with the disaccharide.
J. Mol. Biol., **360**(3), 573-585 (2006).
3. **Takafumi Itoh, Wataru Hashimoto, Bunzo Mikami, and Kousaku Murata:** Crystal structure of unsaturated glucuronyl hydrolase complexed with substrate: molecular insights into its catalytic reaction mechanism.
J. Biol. Chem., **281**(40), 29807-29816 (2006).
4. **Takafumi Itoh, Wataru Hashimoto, Bunzo Mikami, and Kousaku Murata:** Substrate recognition of unsaturated glucuronyl hydrolase from *Bacillus* sp. GL1.
Biochem. Biophys. Res. Commun., **344**(1), 253-262 (2006).
5. **Takafumi Itoh, Akihito Ochiai, Bunzo Mikami, Wataru Hashimoto, and Kousaku Murata:** Structure of unsaturated rhamnogalacturonyl hydrolase complexed with substrate.
Biochem. Biophys. Res. Commun., **347**(4), 1021-1029 (2006).

Related paper

Takafumi Itoh, Bunzo Mikami, Isafumi Maru, Yasuhiro Ohta, Wataru Hashimoto, and Kousaku Murata: Crystal structure of *N*-acyl-D-glucosamine 2-epimerase from porcine kidney at 2.0 Å resolution.
J. Mol. Biol., **303**(5), 733-744 (2000).